# DOCUMENT RESUME

|  |  |
|---|---|
| ED 285 228 | CS 505 672 |

|  |  |
|---|---|
| AUTHOR | O'Brien, Nancy, Ed. |
| TITLE | Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, January-June 1987. |
| INSTITUTION | Haskins Labs., New Haven, Conn. |
| SPONS AGENCY | National Institutes of Health (DHHS), Bethesda, Md.; National Science Foundation, Washington, D.C. |
| REPORT NO | SR-89/90 |
| PUB DATE | 87 |
| CONTRACT | NICHHD-N01-HD-5-2910 |
| GRANT | NICHHD-HD-01994; NIH-BRS-RR-05596; NINCDS-NS-13617; NINCDS-NS-13870; NINCDS-NS-18010; NINCDS-NS-24655; NSF-BNS-8520709 |
| NOTE | 213p. |
| AVAILABLE FROM | U.S. Department of Commerce, National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22151. |
| PUB TYPE | Reports - Research/Technical (143) -- Information Analyses (070) |
| EDRS PRICE | MF01/PC09 Plus Postage. |
| DESCRIPTORS | *Communication Research; Intonation; Lexicology; Linguistics; Morphology (Languages); Phoneme Grapheme Correspondence; *Speech Communication; Speech Habits; Spelling; Vowels; Word Recognition |
| IDENTIFIERS | Vocalization |

ABSTRACT

One of a series of reports on the status of speech investigation, this collection of articles deals with topics including intonation and morphological knowledge. The titles of the articles and their authors are as follows: (1) "Integration and Segregation in Speech Perception" (Bruno H. Repp); (2) "Speech Perception Takes Precedence over Nonspeech Perception" (D. H. Whalen and Alvin M. Liberman); (3) "Evidence of Talker-Independent Information for Vowels" (Robert R. Verbrugge and Brad Rakerd); (4) "Controlled Variables in Sentence Intonation" (Carole E. Gelfer, Katherine S. Harris, and Thomas Baer); (5) "Articulatory Synthesis: Numerical Solution of a Hyperbolic Differential Equation" (Richard S. McGowan); (6) "Type and Number of Violations and the Grammatical Congruency Effect in Lexical Decision" (G. Lukatela, A. Kostic, D. Todorovic, C. Carello, and M. T. Turvey); (7) "Low Constraint Facilitation in Lexical Decision with Single Word Contexts" (G. Lukatela, Claudia Carello, A. Kostic, and M. T. Turvey); (8) "The Use of Morphological Knowledge in Spelling Derived Forms by Learning-Disabled and Normal Students" (Joanne F. Carlisle); (9) "The Development of Morphological Knowledge in Relation to Early Spelling Ability" (Hyla Rubin); (10) The Crossword Puzzle Paradigm: The Effectiveness of Different Word Fragments as Cues for the Retrieval of Words" (Naomi Goldblum and Ram Frost); (11) "On the Possible Role of Auditory Short-Term Adaptation in Perception of the Prevocalic [m]-[n] Contrast" (Bruno H. Repp); (12) "Difference in Second-Formant Transitions between Aspirated and Unaspirated Stop Consonants Preceding [a]" (Bruno H. Repp and Hwei-Bing Lin); (13) "Sensitivity to Inflectional Morphology in Agrammatism: Investigation of a Highly-Inflected Language" (K. Lukatela, S. Crain, and D. Shankweiler); and (14) "Intentionality: a Problem of Multiple Reference Frames, Specificational Information, and Extraordinary Boundary Conditions on Natural Law" (M. T. Turvey). A list of DTIC and ERIC numbers is appended. (HTH)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED285228

# HASKINS LABORATORIES

# STATUS REPORT ON SPEECH RESEARCH

*SR-89/90* *JANUARY-JUNE 1987*

A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for its Investigation, and Practical Applications

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED: The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order number of previous Status Reports.

Haskins Laboratories
270 Crown Street
New Haven, CT 06511

CS 505 672

BEST COPY AVAILABLE

**Michael Studdert-Kennedy, Editor-in-Chief**
**Nancy O'Brien, Editor**
**Yvonne Manning, TEXnician***

---

* This report was produced using **TEX**, the computer typesetting system developed by Donald E. Knuth at Stanford University. **TEX** is a registered trademark of the American Mathematical Society.

The research reported here was made possible
in part by support from the following sources:

NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT

Grant HD-01994

NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT

Contract NO1-HD-5-2910

NATIONAL INSTITUTES OF HEALTH

Biomedical Research Support Grant RR-05596

NATIONAL SCIENCE FOUNDATION

Grant BNS-8520709

NATIONAL INSTITUTE OF NEUROLOGICAL AND COMMUNICATIVE
DISORDERS AND STROKE

Grant NS 13870
Grant NS 13617
Grant NS 18010
Grant NS 24655

## *PERSONNEL IN SPEECH RESEARCH*

### Investigators

Arthur S. Abramson*
Peter J. Alfonso*
Thomas Baer
Fredericka Bell-Berti*
Catherine Best*
Geoffrey Bingham**
Gloria J. Borden*
Susan Brady*
Catherine P. Browman
Etienne Colomb***
Franklin S. Cooper*
Stephen Crain*
Robert Crowder*
Laurie B. Feldman*
Janet Fodor*
Anne Fowler**
Carol A. Fowler*
Ram Frost†
David Garrett‡
Louis Goldstein*
Vicki L. Hanson*
Katherine S. Harris*
Leonard Katz*
J. A. Scott Kelso*
Andrea G. Levitt*
Alvin M. Liberman*
Isabelle Y. Liberman*
Diane Lillo-Martin*
Leigh Lisker*
Anders Löfqvist*
Virginia H. Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard S. McGowan
Kevin G. Munhall
Hiroshi Muta[1]
Patrick W. Nye
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Michael Studdert-Kennedy*
Betty Tuller*
Michael T. Turvey*
Douglas Whalen

### Technical/Support

Philip Chagnon
Alice Dadourian
Michael D'Angelo
Betty J. DeLise
Vincent Gulisano
Donald Hailey
Raymond C. Huey*
Sabina D. Koroluk
Yvonne Manning
Bruce Martin
Vance Maverick
Nancy O'Brien
William P. Scully
Richard S. Sharkany
Edward R. Wiley

### Students*

Joy Armson
Dragana Barac
Eric Bateson
Suzanne Boyce
André Cooper
Margaret Dunn
Carole E. Gelfer
Karen Golding-Kushner
Elizabeth Goodell
Joseph Kalinowski
Bruce Kay
Rena Krakow
Deborah Kuglitsch
Hwei-Bing Lin
Katrina Lukatela
Paul Macaruso
Harriet Magen
Judy Mahoney
Sharon Manuel
Diana Matson
Gerald McRoberts
Sheri L. Mize
Maria Mody
Mark Pitt
Nian-qi Ren
Lawrence D. Rosenblum
Arlyne Russo
Richard C. Schmidt
Jeffrey Shaw
Caroline Smith
Robin Story
Mark Tiede
David Williams
Karen Zaltz

---

*Part-time
**NIH Research Fellow
***Fogarty International Fellow, Lausanne, Switzerland
†Postdoctoral Fellow, Hebrew University, Israel
‡NRSA Training Fellow
[1] Visiting from University of Tokyo, Japan

*CONTENTS*

**Integration and segregation in speech perception**
Bruno H. Repp . . . . . . . . 1

**Speech perception takes precedence over nonspeech perception**
D. H. Whalen and Alvin M. Liberman . . . . . . . . 33

**Evidence of talker-independent information for vowels**
Robert R. Verbrugge and Brad Rakerd . . . . . . . . 39

**Controlled variables in sentence intonation**
Carole E. Gelfer, Katherine S. Harris, and Thomas Baer . . . . . 57

**Articulatory synthesis: Numerical solution of a hyperbolic differential equation**
Richard S. McGowan . . . . . . . . 69

**Type and number of violations and the grammatical congruency effect in lexical decision**
G. Lukatala, A. Kostić, D. Todorović,
C. Carello, and M. T. Turvey . . . . . . . . 81

**Low constraint facilitation in lexical decision with single word contexts**
G. Lukatela, Claudia Carello, A. Kostić, and M. T. Turvey . . . . 93

**The use of morphological knowledge in spelling derived forms by learning-disabled and normal students**
Joanne F. Carlisle . . . . . . . . 105

**The development of morphological knowledge in relation to early spelling ability**
Hyla Rubin . . . . . . . . 121

**The crossword puzzle paradigm: The effectiveness of different word fragments as cues for the retrieval of words**
Naomi Goldblum and Ram Frost . . . . . . . . 133

**On the possible role of auditory short-term adaptation in perception of the prevocalic [m]-[n] contrast**
Bruno H. Repp . . . . . . . . 147

**Difference in second-formant transitions between aspirated and unaspirated stop consonants preceding [*a*]**
Bruno H. Repp and Hwei-Bing Lin . . . . . . . . 173

**Sensitivity to inflectional morphology in agrammatism: Investigation of a highly-inflected language**
K. Lukatela, S. Crain, and D. Shankweiler . . . . . . . . 187

**Intentionality: A problem of multiple reference frames, specificational information, and extraordinary boundary conditions on natural law**
M. T. Turvey . . . . . . . . 199

**Appendix: DTIC and ERIC numbers**
(SR-21/22 - SR-88) . . . . . . . . 207

# *Status Report on Speech Research*

Haskins Laboratories

7

# INTEGRATION AND SEGREGATION IN SPEECH PERCEPTION*

**Bruno H. Repp**

## INTRODUCTION

In this paper I present an overview of some recent research on speech perception. To reduce my task to manageable size, I have chosen to focus on the topics of perceptual integration and segregation, which have guided, more or less explicitly, a considerable amount of speech perception research and theorizing in recent years. This will be a selective review, therefore, but I hope it will nevertheless convey some of the flavor of contemporary ideas and findings, even though that flavor will be tinged with my own favorite spices.

## I. CONCEPTUAL FOUNDATIONS

Integration and segregation are hypothetical perceptual functions (or processes) that link physical structures in the world with mental structures in the brain. An integrative function maps multiple physical units (trivially, a single physical unit) onto a single mental unit, whereas a segregative function maps multiple physical units (sometimes, paradoxically, a single physical unit) onto different mental units. Though mutually exclusive for any particular physical structure at any given time, these two processes nevertheless cooperate in sorting a complex stream of sensory inputs into an orderly sequence of perceived objects and events.

These definitions seem rather straightforward, but they rest on four important assumptions: (1) The physical and mental worlds are not isomorphic. (2) There are objectively definable units in the physical world. (3) There are units in the mental world that are different from the physical units. (4) There are perceptual functions or processes that accomplish the mapping between the two types of units. I will briefly defend each of these assumptions; at the end of this presentation, I will consider the consequences of abandoning some or all of them.

The first assumption, that the mental world is not isomorphic with the physical world, reflects the facts that physical variables are filtered and transformed by sensory systems, that perception is a function not only of the current sensory input but also of the past history of the organism, and that there is often an element of choice in perception that permits alternative perceptual organizations for the same sensory input. Without this assumption, it would be difficult to say anything meaningful about perception, except that it happens.

---

* To appear in *Proceedings of the Eleventh International Congress of Phonetic Sciences*, Tallinn, Estonia, USSR (1987).

*Acknowledgment.* Preparation of this paper was supported by NICHD Grant HD-01994 to Haskins Laboratories. I am grateful to Ignatius Mattingly, Susan Nittrouer, and Michael Studdert-Kennedy for helpful comments on an earlier version.

The second assumption, concerning the existence of physical units, is necessary in order to be able to talk about perceptual integration: These units or dimensions are what is being integrated. Perceptual segregation, too, ordinarily implies that certain objective lines of division can be found in the sensory input. It is always possible to find a physical description that is more finely grained than our description of the perceptual end product. The fact that the machines we use to assess physical characteristics of speech are mere transducers (or, at best, model only peripheral auditory processes) generally assures a mismatch between physical and perceptual descriptions even when the grain size is comparable (and even though our visual perception is engaged in interpreting the machine outputs). Although there are different ways of characterizing the physical energy pattern, they are all equally valid for descriptive purposes. It is an empirical question whether or not perceivers are sensitive to any observed physical divisions, that is, whether these divisions can serve as the basis for perceptual segregation or whether they are bridged by integrative processes. Research of this kind may enable us to find a physical description with a simpler mapping onto perceptual units.

The third assumption concerns the existence and nature of perceptual (mental) units. There is no theory of speech perception that does not assume mental units, usually the ones supplied by linguistic theory. The argument has been over the "perceptual reality" of syllables, phonemes, and features, and over their relative primacy in perceptual processing (see, e.g., Jaeger, 1980; Lehiste, 1972; Massaro, 1975; McNeill & Lindig, 1973; Savin & Bever, 1970). However, which level of the linguistic hierarchy is perceptually and behaviorally salient depends very much on the task and the situation a perceiver is in. As McNeill and Lindig (1973, p. 430) have aptly put it, "what is 'perceptually real' is what one pays attention to." The validity of the basic linguistic categories, questions of detail aside, is guaranteed by the success of linguistic analysis. Linguistic units provide us with a vocabulary in which to describe the time course of accumulation and perceptual processing of linguistic information. Even though the perceptual processes themselves may be of an analog nature, we need discrete concepts to theorize and communicate about these processes. From this perspective, it is not an empirical issue but a fact that perceivers process features, phonemes, syllables, words, etc., since they are what speech is made of. Their *awareness* of these categories is another matter that shall not concern us here. (See Mann, 1986; Mattingly, 1972; Morais, Cary, Alegria, & Bertelson, 1979.) Clearly, speech perception generally proceeds without awareness of all but the highest levels of description (i.e., the meaning of the message).

The fourth assumption is that there are perceptual processes in the brain that map sensory inputs onto internal structures. While such processes have been traditionally assumed in psychology since the demise of radical behaviorism, a new challenge (to the other assumptions as well) comes from the so-called direct realist school of perception, which claims that perceptual systems merely "pick up" the information delivered by the senses (Fowler, 1986; Gibson, 1966). I will return to this issue later. Here I merely note that the same input is not always perceived in the same way. Contextual factors, past experience, expectations, and strategies may alter the perceptual outcome, and this seems to require the assumption of perceptual processes that mediate between the input and the perceiver's interpretation of it. Whether these processes (and indeed, integration and segregation as such) are thought of as neural events with actual time and space coordinates or as abstract functional relationships between physical and mental descriptions is irrelevant to most of the research I will discuss here.

Having attempted to justify the four principal assumptions, it remains for me to mention two issues that are important in much research on perceptual integration and segregation. One is the question of whether the processes inferred are specific to the perception of speech or whether they represent general capacities of the auditory or cognitive system. By a speech-specific function I mean one that operates on properties that are unique to speech. There is no question that general capacities to integrate and segregate are common to all perceptual and cognitive systems. Speech perception presumably results from a combination of general and speech-specific perceptual functions (see, e.g., Diehl, 1987), just as speech resembles other sounds in some respects and differs in others. One frequent research strategy, therefore, is to determine whether or not *particular* instances of integration or segregation can be observed in both speech and nonspeech perception. This question can be asked only if the physical characteristics of speech and nonspeech stimuli are comparable—a condition that is notoriously difficult to satisfy (see, e.g., Pisoni, 1987). The mental descriptions of speech and nonspeech are, by definition, different at some higher level; thus the empirical question is whether that level is engaged in a particular integrative or segregative process.

The other issue is whether a particular integrative or segregative function is obligatory or optional. This question is sometimes linked with that of speech-specificity in that a higher-level, speech-specific function might seem easier to disengage than a lower-level auditory one. This is true in so far as adopting the deliberate strategy of listening to speech as if it were nonspeech (which is often difficult to achieve) may have the effect of eliminating certain forms of integration or segregation. It seems to be difficult or impossible to disengage phonetic processes through conscious strategies within the speech mode (e.g., by linguistic parsing; Repp, 1985a, 1985b). Moreover, it has been suggested (Liberman & Mattingly, 1985) that some speech-specific functions do not really represent a "higher" level of perception but rather a mode of operation that, because of its biological significance, takes precedence over nonspeech perception, and if so, these functions may indeed be difficult to manipulate. On the other hand, in the auditory (nonspeech) mode listeners often have a variety of perceptual strategies available, especially when there are few ecological constraints on the stimulation, even though certain functions of peripheral auditory processing are surely obligatory. Thus, although it is useful to gather information about the relative flexibility of a process, this may not bear directly on the question of speech-specificity, as both speech and nonspeech perception are likely to involve levels of varying rigidity.

One final prefatory remark: Although one may legitimately talk about the integration of syllables into words and of words into sentences, or about the segregation of syntactic constituents from each other, I am not going to consider such higher linguistic processes in the present review. By speech perception I mean primarily the perception of phonetic structure without regard to lexical status or meaning, and my review is restricted accordingly.

## II. INTEGRATION

The function of integrative processes is to provide coherence among parts of the input that "belong together" according to some perceptual rule or criterion. Auditory integration occurs within the physical dimensions of time, (spectral) frequency, and even space (in the case of artificially split sources); thus it creates temporal, spectral, and spatial coherence of sound sources. In part this is due to the limited resolution of the auditory system along each of these dimensions, but auditory events will often cohere even when there are discriminable changes within them. The larger these changes are, the more noteworthy the integrative process will seem to us. The perception of phonetic structure involves, in addition, integration of relevant information across all physical dimensions of the speech signal—a function requiring higher-level perceptual or cognitive mechanisms.

## A. Temporal Integration

Basic processes of sensory integration and auditory organization ensure the temporal coherence of any relatively homogeneous auditory input, including components of speech. This form of integration is so obvious as to hardly deserve comment. Thus, for example, successive pitch periods of a vowel are perceived as belonging together (i.e., as a single vowel, not two or many) even though their duration and spectral composition may change as a function of intonation, diphthongization, and coarticulation. While there may be a physical basis for subdividing a sound into smaller units such as individual pitch pulses or transition versus steady state, the rate and extent of change from one unit to the next are too small to disrupt sensory integration. Nevertheless, changes occurring within such units (e.g., transitions in a vowel or fricative noise) may have perceptual effects. That is, perception of temporal coherence does not imply insensitivity to changes over time, only that these changes are not large enough to cause perceptual segregation.

### *1. Growth of Loudness*

Temporal integration at this most elementary level has the consequence that, as the duration of a relatively homogeneous sound increases, its perceived loudness or perceptual prominence will also increase, up to a certain limit. In psychoacoustic research, the lowering of the detection threshold and the growth of loudness with increasing stimulus duration are well-established phenomena (see, e.g., Cowan, in press; Zwislocki, 1969). The time constant of the (exponential) integration function is about 200 ms, which encompasses the durations of virtually all relatively homogeneous speech events. While loudness judgments or explicit threshold measurements are uncommon in speech perception research, the effect of an increase in the duration of a signal portion can be shown to be phonetically equivalent to that of an increase in its intensity, especially when the relevant signal portion is brief.

One example is provided by studies in which the duration and relative intensity of aspiration noise were varied orthogonally as cues to the voicing distinction in synthetic syllable-initial English stop consonants (Darwin & Seton, 1983; Repp, 1979b). Although the trading function obtained was much steeper than the typical auditory temporal integration function, it bore some similarity to integration functions obtained in an auditory backward masking situation (Wright, 1964), which is not unreasonable in view of the following vowel. It seems likely that the observed time-intensity reciprocity reflects basic properties of the auditory system, rather than speech-specific processes. Indirect support for this hypothesis comes from a study showing that the trading relation between aspiration duration and intensity holds regardless of whether or not listeners can rely on phonemic distinctions in discriminating speech stimuli (Repp, 1983b). In another recent study, stop consonant release burst duration and intensity were varied in separate experiments as cues to stop consonant manner in /s/-stop clusters (Repp, 1984c). Since both parameters proved to be perceptually relevant, a trading relation between them was implied. An analogous conclusion may be drawn from an older informal study by Lisker (1978), in which the duration and intensity of stop closure voicing were varied as cues to the perceived voicing status of an intervocalic stop consonant.

### *2. Auditory Short-term Adaptation*

An effect closely related to temporal integration is that the auditory nerve fibers responsive to a continuous sound become increasingly adapted. Auditory adaptation is a topic of great interest to psychoacousticians and auditory physiologists, who have identified at least three different time constants of adaptation in animals (see, e.g., Eggermont, 1985). So-called auditory short-term adaptation, with a time constant of about 60 ms, seems the most relevant to phonetic perception. Although ongoing adaptation seems to have no direct perceptual consequences, the recovery of auditory nerve fibers following the offset of a relatively homogeneous stimulus results in reduced sensitivity to other, spectrally similar inputs for a short time period. Consequently, the auditory representation of a speech component whose spectrum overlaps that of a preceding segment will be modified. A striking demonstration of such an interaction was provided by Delgutte (1980; Delgutte & Kiang, 1984) in recordings from cats' auditory nerves responding to synthetic /ba/ and /ma/ syllables. Even though the two syllables were identical except for the nasal murmur in /ma/, the auditory response at vowel onset was very different. The murmur, having strong spectral components in the low-frequency range, effectively acted as a high-pass filter, reducing the neural response in the low-frequency region at vowel onset. Recent experiments suggest, however, that this particular auditory interaction has no important consequences for perception of nasal consonants under normal listening conditions (Repp, 1987a). In a more artificial situation, Summerfield, Haggard, Foster, and Gray (1984) and Summerfield and Assmann (1987) have demonstrated an auditory aftereffect attributed to short-term adaptation: A sound with a uniform spectrum was perceived as a vowel when preceded by a sound whose spectrum was the complement of the perceived vowel's spectrum. Generalizing to natural speech, these authors pointed out that auditory adaptation effectively enhances spectral change and thus may aid phonetic perception in adverse listening conditions.

One general lesson to be learned from psychoacoustic research on temporal integration, adaptation, and other auditory interactions is that adjacent portions of the speech signal should not be thought of as mutually independent in the auditory system. Whenever a particular component is singled out for attention in careful analytic listening (to the extent that this is possible), influences of surrounding context on the perceived sound must be reckoned with. It is important to keep in mind, however, that listeners normally do not listen analytically but rather attend to the continuous pattern of speech. All peripheral auditory transformations are a natural part of the pattern and, because of past learning, are also represented in a listener's long-term memory of phonetic norms, which provide the criteria for phonemic classification in a language. Since auditory input and central reference both incorporate the distortions imposed by the peripheral auditory system, these distortions cannot be said to either help or hinder speech perception (see Repp, 1987b). Only a change in auditory transformations, as might be caused by simulated or real hearing impairment, would prove disturbing to listeners; in normal speech perception, peripheral auditory processes probably do not play a very important role.

## B. Spectral Integration

Most speech sounds have complex spectra determined by the resonance frequencies of the vocal tract. Formants are usually visible as prominent energy bands in a spectrogram or as peaks in a spectral cross-section. Why are these bands perceived as a single sound with a complex timbre and not as separate sounds with simpler qualities? Why, indeed, are the individual harmonics of periodic speech sounds not heard as so many simultaneous tones? Even though these questions are provoked by our instrumental and visual methods of spectral analysis, they are not unreasonable, since the ear operates essentially as a frequency analyzer. One answer to these questions is that we *do* process these spectral components, only we are not conscious of them and find it difficult to focus selectively on them when asked to do so. Multidimensional statistical analyses of vowel similarity judgments have confirmed that the lower formants function as perceptually relevant dimensions, even though they seem to blend into a complex auditory quality (e.g., Fox, 1983; Pols, van der Kamp, & Plomp, 1969; Rakerd & Verbrugge, 1985), and psychoacoustic pitch matching tasks have revealed that listeners can detect a number of lower harmonics in a complex periodic sound (e.g., Peters, Moore, & Glasberg, 1983; Plomp, 1964). Some central integrative function must be responsible for the perceptual coherence and unity of all these spectral components.

### *1. Critical Bands*

Some spectral integration does take place in the peripheral auditory system. A large amount of psychoacoustic research has established the concept of critical bands, i.e., frequency regions over which spectral energy is integrated, and whose width increases with frequency in a roughly logarithmic fashion (Moore & Glasberg, 1983; Zwicker & Terhardt, 1980). It is now quite common to represent speech spectra on a critical-band frequency scale (the Bark scale) to better take account of the resolving power of the auditory system. However, critical bands cannot account for the fact that formants are integrated into a unitary percept, because the lower formants of speech are usually several critical bands apart, and thus potentially separable. Even the lower harmonics, especially of female and child speech, are spaced more than 1 Bark apart. Critical bands may explain why higher harmonics and higher formants are not well resolved auditorily, but these spectral components do not contribute much phonetic information.

It is difficult, therefore, to point to any direct consequences of critical band limitations for speech perception, except in hearing-impaired listeners, whose critical bandwidths are abnormally large. A recent study by Celmer and Bienvenue (1987) may serve as an example. These investigators digitized speech materials, degraded their spectra by simulating critical band integration ranging from one-half to seven times the normal widths, converted the manipulated spectra back into sound, and presented them to groups of normal listeners and to hearing-impaired listeners believed to have abnormally wide critical bandwidths according to independent psychoacoustic tests. The results showed that the degree of critical bandwidth filtering required to cause an intelligibility decrement was directly related to the subjects' measured critical bandwidth. Thus, normal subjects were sensitive to filtering at twice the normal bandwidths, while hearing-impaired subjects, though their intelligibility scores were lower to begin with, tolerated up to five times the normal bandwidths before any decrement in intelligibility occurred. Many other studies, too numerous to review here, have examined correlations between measures of critical bandwidth (or frequency resolution) and measures of speech perception in hearing-impaired individuals, with mixed results (see, e.g., Dreschler & Plomp, 1980; Stelmachowicz, Jesteadt, Gorga, & Mott, 1985). The looseness of the correlation may be accounted for by the facts that speech perception engages higher-level functions that help overcome peripheral limitations, often requires only relatively coarse spectral resolution, and relies on other physical parameters besides spectral structure.

## *2. Integration of Harmonics*

Given that the lower harmonics of a periodic speech sound are not automatically integrated by the peripheral auditory system, not to mention the lower formants themselves, the question of why they are grouped together in perception still needs to be answered. The most general answer is that they share a "common fate": They usually start and end at the same time; they are at integral multiples of the fundamental frequency; they have similar amplitude envelopes; and there is no alternative grouping that suggests itself. Below I will have more to say about the factors that may cause segregation of harmonics. Principles of auditory organization have received much attention in recent years (see, e.g., Bregman, 1978; Darwin, 1981; Weintraub, 1987), and one interesting conclusion from that research is that, even at such a relatively early stage in auditory processing, speech-specific criteria begin to play a role. They are speech-specific in the sense that a listener's tacit knowledge of what makes a good speech pattern influences the perceptual grouping of auditory components, as presumably does knowledge of other familiar auditory patterns. Yet another answer to the question of why harmonics (and formants) are grouped together is, therefore: They make a speech sound—that is, a complex sound that could possibly have emanated from a human vocal tract.

If it is the case that formant frequencies are salient parameters of speech perception (an assumption that is not made by some researchers who favor a whole-spectrum approach; e.g., Bladon, 1982; Stevens & Blumstein, 1981), then it is of interest to ask how listeners estimate the actual resonance frequencies of the vocal tract from the energy distribution in the relevant spectral region. This question is especially pertinent with respect to the first formant ($F_1$) in periodic speech sounds, for which critical bands are narrow and frequency difference limens are small. This means that the actual $F_1$ frequency often falls between auditorily resolvable harmonics. Early work by Mushnikov and Chistovich (1973) suggested that the brain takes the frequency of the single most intense harmonic as the estimate of $F_1$. Later studies by Carlson, Fant, and Granström (1975) and Assmann and Nearey (1987), however, have indicated that the subjective $F_1$ frequency corresponds to a weighted average of the two most intense harmonics, and Darwin and Gardner (1985) have shown that the perceptual boundary between /ɪ/ and /e/ can be affected by the intensity of as many as five harmonics between 250 and 750 Hz, spaced 125 Hz apart. This indicates that the weighting function applied by the speech perception system in estimating formant frequencies extends over several critical bands (which are 100 Hz or less in this frequency region). The function is also asymmetric, giving more weight to higher than to lower harmonics, which may reflect a speech-specific constraint related to the fact that changes in actual $F_1$ frequency affect primarily the amplitudes of the higher harmonics in the vicinity of the spectral peak (Assmann & Nearey, 1987). Listeners thus seem to have tacit knowledge of the physical constraints on the shape of the vocal tract transfer function (Darwin, 1984).

## *3. Integration of Formants*

This leads us to the more general question of whether the speech perception system integrates over adjacent formants (or any two peaks in the spectrum) when they are close in frequency but not within a critical band. It has been known for a long time that reasonable approximations to virtually all vowels can be achieved in synthesis with just two formants, and even with a single formant in the case of back vowels (Delattre, Liberman, Cooper, & Gerstman, 1952). Delattre et al. noted that the approximations were best when the two formants replaced by a single formant were close in frequency ($F_1$ and $F_2$ in high back vowels; $F_2$ and $F_3$ in high front vowels), and that the best single-formant substitute tended to be intermediate in frequency, suggesting that closely adjacent vowel formants form a perceptual composite or average. This idea was later elaborated by the Stockholm research group (Carlson, Granström, & Fant, 1970; Carlson et al., 1975) into the concept of $F_2'$, a hypothetical effective formant intermediate in frequency between $F_2$ and $F_3$ (except for /i/, where it falls between $F_3$ and $F_4$). These authors developed a formula for calculating $F_2'$ from $F_1$, $F_2$, $F_3$, and $F_4$, which gave good approximations to the results of perceptual matching experiments.

More recently, Chistovich and her collaborators have conducted a number of experiments on the "center of gravity" effect—the demonstrable phonetic equivalence of a single formant to two adjacent formants of varying frequency and/or intensity (see Chistovich, 1985, for a review). One important question concerned the critical frequency separation of the two formants beyond which no satisfactory single-formant match could be achieved; it turned out to be about 3.5 Bark, that is, 3.5 critical bands (Chistovich & Lublinskaja, 1979). This finding has received considerable attention. For example, the 3.5 Bark limit has been related to the separation and boundaries between English vowel categories in acoustic space (Syrdal & Gopal, 1986), and it has been used, together with the center of gravity concept, to explain perceived shifts in the height of nasalized vowels, which often have two spectral prominences in the $F_1$ region (Beddor, 1984).

It is noteworthy, however, that already Delattre et al. (1952) were unable to achieve satisfactory single-formant matches to arbitrary two-formant patterns that did not correspond to familiar vowel categories. This finding, which was replicated by Traunmüller (1982, 1984b), suggests that spectral integration over 3.5 Bark is tied to the perception of phonetic (or phonemic) categories. Specifically, it may reflect the resolution of the auditory long-term memory in which phonetic reference patterns are stored (Traunmüller, 1984b). Indeed, it is an open question whether the 3.5 Bark limit explains the acoustic spacing of vowel categories (Syrdal & Gopal, 1986), or whether it is the other way around. A recent study by Schwartz and Escudier (1987), however, provides evidence that the 3.5 Bark limit is not the consequence of phonemic categorization. Their data suggest that there is indeed a higher level of auditory representation that serves phonetic classification and includes wide-band spectral integration. The cause of this integration is unknown at present.

### *4. Redintegration of Artificially Separated Spectral Components*

Ultimately, it must be a higher-level process that decides whether a spectral array constitutes a single event or several. Integration over the whole spectrum is the natural state of affairs, since most natural sounds have complex spectra and could not easily be recognized if integration were not the default operation. Even an unrelated set of pure tones is perceived as a single complex structure when sounded simultaneously, as long as no alternative organizations suggest themselves (e.g., Green, 1983; Kubovy, 1981). Such integration is disrupted by temporal or spatial separation of signal components, however; for example, the "auditory profiles" studied by Green and his coworkers are not well perceived when the sinusoidal components are divided between the two earphone channels (Green & Kidd, 1983). With familiar natural events such as speech, perceptual coherence of spectral components may be centrally guided and hence greater and more resistant to disruption. One possible example of this is the phenomenon called spectral-temporal fusion (Cutting, 1976) or duplex perception (Liberman, 1979), which has been studied extensively in recent years.

Precursors of this research are found in experiments where the formants of synthetic syllables were separated and presented to opposite ears (e.g., $F_1$ to one ear and $F_2$ and $F_3$ to the other). It was found early on that this presentation gave rise to an intact speech percept, with little or no awareness of separate stimuli in the two ears (Broadbent & Ladefoged, 1957). Similar fusion of dichotic stimuli into a single perceived sound is observed with complete synthetic syllables in the two ears (e.g., Repp, 1976b) and even with harmonically related tones (e.g., Deutsch, 1978). More surprising is the finding that perceptual integration continues to occur even when listeners are aware of separate stimuli in the two ears. Thus, Cutting (1976) presented the dichotically separated formants at different fundamental frequencies and observed that subjects still reported the percept corresponding to the combination of the formants. (For similar effects with diotic presentation, see Darwin, 1981.) In what is now called the duplex perception paradigm, Rand (1974) presented the formant transitions distinguishing two synthetic consonant-vowel syllables (such as /da/ and /ga/) to one ear and the remainder common to the two syllables (the "base") to the opposite ear. In this situation, listeners continue to report one or the other syllable depending on which formant transition is presented, even though that transition is also heard simultaneously as a lateralized nonspeech "chirp." The intact syllable (not the base) is heard in the ear receiving the base. Thus, subjectively at least, auditory fusion takes place despite the auditory segregation of the chirp—a paradoxical situation. This fusion continues to operate when the two signal components are presented at different fundamental frequencies (Cutting, 1976) or with slight temporal offsets (Repp & Bentin, 1984). A very similar phenomenon can be produced diotically by making the critical formant transition audible through temporal offset (Repp & Bentin, 1984), amplification (Whalen & Liberman, 1987), or different fundamental frequencies (informal observations). None of these manipulations, within certain limits, destroys the fused speech percept.

One interpretation of these findings (see, e.g., Liberman & Mattingly, 1985) is that a specialized speech "module" is responsible for the perceptual integration and apparent fusion, whereas the general auditory system is responsible for the separate chirp percept. Bregman (1987), on the other hand, has proposed that the paradoxical co-occurrence of fusion and nonfusion arises from conflicting cues for integration and segregation in the general process of "auditory scene analysis." He and other students of auditory organization have stressed the relative independence of *What* and *Where* decisions in auditory perception (Bregman & Steiger, 1980; Darwin, 1981; Deutsch & Roll, 1976; Weintraub, 1987). It seems that auditory components that have been segregated can nevertheless be recombined in the perception and classification of familiar sound structures. That this recombination in the duplex perception paradigm is genuinely perceptual and not cognitive is indicated not only by the subjective impression of an intact syllable but by the fact that the components (chirp and base) presented by themselves generally do not suggest the "correct" phonetic percept (Repp, Milburn, & Ashkenas, 1983).

## C. Integration of Phonetic Information

Speech consists of a sequence of diverse sound segments that, as everyone knows, do not correspond directly to linguistic units. Changes in spectral structure are often very rapid and lead to great spectral heterogeneity over time. Equally striking is the alternation of qualitatively different sound types (periodic vs. aperiodic, as well as silence). Nevertheless, listeners perceive a coherent event, and thus believe speech to be a coherent stream of sounds. Since there is absolutely no reason to assume that very disparate sound structures are automatically integrated by the auditory system, the subjective impression of auditory continuity must be due to higher-level articulatory and linguistic properties of cohesiveness that capture the listener's attention—a kind of categorical perception (see Repp, 1984a).

How can our brain perform integrative feats in speech perception that exceed the capabilities of the auditory system? One possibility is that there exists a biological specialization in humans, a "speech module," which performs this task (see Fodor, 1983; Liberman & Mattingly, 1985). Alternatively, the answer may be mental *precompilation* as a consequence of perceptual learning—an assembled module, as it were (cf. Klatt, 1979). What distinguishes speech perception from the auditory perception of arbitrary tones and noises (but not necessarily from the perception of other ecologically significant auditory events) is that the input can be mapped onto meaningful units of various sizes. The integration of the auditory components relating to each unit represented in the perceiver's long-term memory has taken place long ago during the process of speech and language acquisition; it may be instantiated neurally as a flexible (context-sensitive) system of interconnections (Elman & McClelland, 1984; Klatt, 1979). These precompiled units then enable a perceiver to immediately relate a number of functionally independent auditory features to a common phonetic percept. Some interesting (and arduous) attempts to simulate this process of perceptual learning and unit formation in nonspeech auditory perception have been reviewed by Watson and Foyle (1985), who stress the importance of central processes in the identification and discrimination of complex stimuli. Experienced Morse code operators exhibit similar skills of "integrating" the acoustic dots and dashes into larger units (Bryan & Harter, 1899), and so do probably perceivers of other meaningful acoustic events in our environment (see Jenkins, 1985; Warren & Verbrugge, 1984), although in none of these instances does the auditory stimulus structure recede as much from awareness as it does in speech perception. From this perspective, speech is unique not so much because it requires specialized perceptual and cognitive functions but because it is structurally different, having originated in the articulatory motor system. Our biological specialization may simply lie in the fact that we can mentally represent a system that complex.

### 1. "Integrated" Auditory Properties

The ability to integrate over dynamically changing sound patterns has occasionally been attributed to the auditory system. Thus, Stevens and Blumstein (1978, 1981; Blumstein & Stevens, 1980) hypothesized that the onset spectrum following the release of stop consonants provides invariant acoustic correlates of place of articulation. Since there are often rapid spectral changes immediately following the release, and since a spectrum cannot be computed instantaneously, the hypothetical auditory onset spectrum must derive from an integrative process. Stevens and Blumstein hypothesized that the human auditory system integrates over about 25 ms and thus extracts the acoustic property relevant to place of articulation.

The work of Stevens and Blumstein has come under criticism in recent years. Kewley-Port (1983) has argued that, for all we know, the auditory system tracks spectral changes over time intervals shorter than 25 ms and presumably delivers information about these changes to phonetic decision mechanisms. A perceptual study by Kewley-Port, Pisoni, and Studdert-Kennedy (1983) has suggested that listeners are indeed sensitive to spectral changes immediately following the release of stop consonants (see also Blumstein & Stevens, 1980). The onset spectra themselves do not appear to be as invariant as was originally claimed (see Lahiri, Gewirth, & Blumstein, 1984; Suomi, 1985). Blumstein and her students meanwhile have abandoned the search for invariant properties in onset spectra and have instead gone on to define integrated properties based on the relationship between spectra or intensity measures obtained some interval apart (Jongman, Blumstein, & Lahiri, 1985; Kurowski & Blumstein, in press; Lahiri et al., 1984). Even though some of these properties are quite complex, their derivation is still attributed to the auditory system by these researchers. However, since it seems highly implausible that there are general auditory functions that yield so specialized a result, the epithet "auditory" should perhaps be understood as referring merely to the input modality. Clearly, out of the infinity of possibilities, particular relational properties are selected on the basis of phonetic relevance. The integrative computational process thus is specific to speech perception.

## *2. Integration of Silence and Other Signal Components*

Even though it seems unlikely that the auditory system integrates over spectral variation in the speech signal lasting tens of milliseconds, this hypothesis has some measure of plausibility, given the basic continuity of the signal changes. There are many more abrupt changes in the speech signal, however, such as changes in source (from voiced to voiceless, or vice versa), in spectrum (such as /z/ followed by /u/), and in intensity (into and out of closures filled with nasal murmur, voicing, or silence), usually in several of these dimensions simultaneously. It would seem absurd to attribute to the auditory system the capability to integrate across such dramatic signal changes, since the task of auditory perception is to detect changes, not to conceal them. Nevertheless, there is ample evidence from perceptual experiments that listeners can integrate phonetic information across such acoustic discontinuities in the signal. Clearly, this integration must be a higher-level function in the service of speech perception.

Perhaps the most striking instance is the perception of silence in speech. (I have in mind brief silent intervals of up to 200 ms duration, not longer pauses.) From an auditory perspective, silence is the absence of energy, a gap, an interruption that separates the signal portions to be perceived. In speech perception, however, silence is bridged by, and participates in, integrative processes. Rather than being the neutral backdrop for the theater of auditory events, silence is informationally equivalent to energy-carrying signal portions. Relative duration of silence has been shown to be a cue for the perception of stop consonant voicing (Kohler, 1979; Lisker, 1957; Port, 1979), manner (Bailey & Summerfield, 1980; Repp, 1984c; Repp, Liberman, Eccardt, & Pesetsky, 1978), and place of articulation (Bailey & Summerfield, 1980; Port, 1979; Repp, 1984b). Why does silence function in this way in speech? The answer must be that it is an integral part of the acoustic patterns that a human listener has learned to recognize. Being an acoustic consequence of the oral closure connected with (voiceless) stop consonants, it has become a defining characteristic of that manner class. Lawful variations in its duration as a function of voicing status or place of articulation also have assumed the function of perceptual "cues." A listener's long-term representation of the acoustic pattern corresponding to a stop consonant thus includes the spectro-temporal properties of the signals preceding and following the closure as well as the closure itself. (The precise nature of that mental representation, or rather of our description of it, need not concern us here; it suffices to note that listeners behave *as if* they knew what acoustic pattern to expect.) The silence thus is not really "actively" integrated with the surrounding signal portions; rather, the integration has already taken place during past perceptual learning and is embodied in the perceiver's long-term knowledge of speech patterns to which the input is referred during perception.

Not only is silence integrated (in the sense just discussed) with surrounding signal portions in phonetic perception, but acoustically rather different components of the signal are integrated with each other. Thus, for example, the spectrum of a fricative noise and the adjacent vocalic formant transitions both contribute to perception of a prevocalic fricative consonant (e.g., Mann & Repp, 1980; Whalen, 1981), the formant transitions in and out of a closure contribute to stop consonant perception (Tartter, Kat, Samuel, & Repp, 1983), etc. Just as articulation distributes acoustic information about individual phonemes over time, perceptual integrative functions collect that information and relate it to internal criteria for linguistic category membership. An especially interesting demonstration of this was provided quite recently by Tomiak, Mullennix, and Sawusch (1987). Using a well-known technique (Garner, 1974) for testing listeners' ability to selectively attend to stimulus dimensions, they showed that the "fricative noise" and "vowel" portions of noise-tone analogs to fricative-vowel syllables were processed separately by subjects who perceived the stimuli as nonspeech sounds, but were processed integrally by subjects who had been told that the stimuli represented syllables. These latter subjects were unable to selectively attend to either of the two stimulus portions, even though coarticulatory interactions were not present in the noise-tone stimuli. Listeners in the "speech mode" thus seem to process auditory components of speech in an integrative manner even if some of the information to be integrated is not actually there; they are scanning for it, as it were.

Independent aspects of the speech signal that contribute to the same phonemic decision combine according to a simple decision rule, as demonstrated in many experiments by Massaro (e.g., Derr & Massaro, 1980; Massaro & Oden, 1980). It is possible to trade various of these cues, changing the physical parameters of one while changing those of another in the opposite direction, without altering the phonemic percept. This phenomenon, often referred to as "phonetic trading relations," has been demonstrated in a large number of studies (see review by Repp, 1982). Fitch, Halwes, Erickson, and Liberman (1980) showed that listeners have great difficulty discriminating two phonemically equivalent stimuli created by playing off two cues against each other, and they argued that this reflects the operation of a special phonetic process that makes auditory differences unavailable to perception. Whether the *process* of phonetic information integration is speech-specific is debatable (cf. Repp, 1987b), even though it is agreed that the information being integrated is speech-specific. Listeners' difficulty in discriminating phonemically equivalent stimuli is familiar from classical categorical perception research (see review by Repp, 1984a). Experiments on phonetic trading relations that include identification and discrimination tests (Best, Morrongiello, & Robson, 1981; Fitch et al., 1980) are generalized categorical perception tasks, in which several physical parameters are varied simultaneously. If each parameter variation by itself is difficult to discriminate except when it cues a category distinction, then joint variations in these parameters will be almost as difficult to discriminate unless a phonemic contrast is perceived. This does not mean, however, that auditory discrimination of such variations is impossible. Appropriate training and use of low-uncertainty discrimination paradigms has been shown to reduce or eliminate categorical perception of single dimensions (Carney, Widin, & Viemeister, 1977; Repp, 1981), and it is likely that similar training would enable subjects to discriminate simultaneous variations in several cues, thus demonstrating that their integration does not take place in the auditory system (see also Best et al., 1981). There is also evidence that certain phonetic trading relations occur *only* when listeners can make phonemic distinctions, but not within phonemic categories (Repp, 1983b).

In summary, the various forms of phonetic cue integration seem to represent, for the most part, speech-specific functions in so far as the articulatory processes and the corresponding linguistic categories that cause the integration are specific to speech. This idea is embodied in Massaro's "fuzzy logical model" of phonetic decision making (Massaro & Oden, 1980), which assumes that, for each phonemic category, listeners have internal criteria for the degree of presence of various acoustic features in the speech signal. Diehl and his colleagues have recently argued that many trading relations may have a general auditory basis (Diehl, 1987; Parker, Diehl, & Kluender, 1986). While their research may show that some trading relations (especially those within a physical dimension) indeed rest on auditory interactions, this is unlikely to be true for the many trading relations that cut across physical dimensions. Although phonetic perception is certainly not immune to auditory interactions, cue integration appears to be mainly a function of speech-specific classification criteria.

### *3. Phonetic Context Effects*

Perceivers not only integrate cues directly pertaining to a particular phoneme or complex of articulatory gestures, but they adapt their perceptual criteria to the surrounding phonetic context. Examples of such phonetic *context effects* are the shift in the /s/-/ʃ/ category boundary depending on the following vowel (Kunisaki & Fujisaki, 1977; Mann & Repp, 1980) and the shift in the /b/-/p/ voice-onset-time category boundary depending on the speaking rate or duration of the surrounding segments (Green & Miller, 1985; Miller, 1981; Summerfield, 1981). For reviews, see Miller (1981), Repp (1982), and Repp and Liberman (1987). As in the case of phonetic trading relations, some of these effects may have general auditory processing explanations; thus, for example, the effect of vowel duration on perception of the /ba/-/wa/ distinction (Miller & Liberman, 1979) probably is not speech-specific, as a comparable effect has also been obtained with nonspeech stimuli (Pisoni, Carrell, & Gans, 1983). Many other effects, however, seem to reflect listeners' tacit knowledge of coarticulatory dependencies in speech production. For example, the different /s/-/ʃ/ boundaries in the context of rounded and unrounded vowels may be related to the occurrence of anticipatory liprounding during the constriction phase in utterances such as "soup" but not in "sap." In a series of experiments using cross-spliced fricative noises and vowels, Whalen (1984; Whalen & Samuel, 1985) has shown that even when the fricative noise itself is quite unambiguous, subjects' reaction time in a fricative identification task is influenced by the following vocalic context, being slower when the fricative noise spectrum is not exactly what would be expected in that context (cf. the study by Tomiak et al., 1987, reviewed above). In an unpublished series of experiments, Repp (1978a) demonstrated an effect he dubbed "coperception," which consisted of slower reaction times to decide that the two consonants are the same in the stimulus pair /aba/-/abi/ than in the pair /aba/-/aba/, even though the pre-closure (VC) portions of these synthetic VCV stimuli were identical in both cases. That is, even though subjects could have made their decisions after hearing /ab/ in the second member of a stimulus pair, they somehow had to take the CV portions of the stimuli into account and then were slowed down by the inequality of the vowels. All these studies show that perceivers integrate all information that possibly could bear on phonetic decisions, and this integration often seems obligatory in nature. It requires special instructions, special (nonphonetic) tasks, and usually some amount of training to disengage phonetic integration mechanisms in the laboratory (e.g., Best et al., 1981; Repp, 1980, 1981, 1985b).

### *4. Cross-modal Integration*

In natural speech communication, humans make use not only of auditory but also of visual information, if available. Audiovisual integration at the level of phoneme perception has been a research topic of considerable interest since the discovery by McGurk and MacDonald (1976) that subjects presented with certain conflicting auditory and visual speech stimuli report that they "hear" what they see. Their findings have been replicated and extended in a number of studies (MacDonald & McGurk, 1978; Massaro & Cohen, 1983; Summerfield, 1981; and others). Massaro (in press; Massaro & Cohen, 1983) has shown that a general rule of information integration based on the degree to which signal features match expected feature values can explain audiovisual integration, auditory cue integration, as well as many other forms of perceptual integration outside the domain of speech. This suggests that we may be dealing with a general function following basic laws of decision theory. Liberman and his collaborators (Liberman, 1982; Repp et al., 1978), on the other hand, have argued that integration of speech cues, within or across modalities, occurs because they represent the multiple, distributed consequences of articulatory acts or gestures. Some internal reference to processes of speech production is thus implied, as in the "motor theory" of speech perception (see Liberman & Mattingly, 1985). However, this account is complementary rather than antithetic to Massaro's model: It is a theory of *why* integration occurs, whereas Massaro is concerned with *how* integration works. The phonemes of a language are articulatory events that have characteristic acoustic and optic consequences, and perceivers presumably have tacit knowledge incorporating both of these aspects. If a portion of the speech input satisfies certain auditory and visual criteria for phonemic category membership (as in Massaro's model) this also implies that the gestures characterizing a particular phoneme have been recovered (as in the motor theory). Whether the sensory or the articulatory aspect is stressed in a particular theory is largely a matter of philosophy and perhaps of economy. A complete theory must include both.

Audiovisual integration at the more global level of word, sentence, and discourse comprehension has, of course, been of interest for a long time in connection with hearing impairment and communication in noisy environments. Research on this topic has received a boost in recent years with the advent of modern signal processing technology and of cochlear implants. (See Summerfield, 1983, for a review.) The information provided by residual hearing or by electrical stimulation of the auditory nerve supplements that obtained from lipreading to yield enhanced comprehension. In many respects, these two sources of information are complementary, with the auditory channel providing information that is difficult to see, and vice versa. What is of special interest in the present context is that audiovisual comprehension performance often seems to exceed what might be expected from a mere combination of independent sources of information. Thus, Rosen, Fourcin, and Moore (1981) demonstrated that speech intelligibility is improved substantially when lipreading in hearing subjects is supplemented with the audible fundamental frequency contour, or even just with a constant buzz representing the occurrence of voicing. (See also Breeuwer & Plomp, 1986; Grant, Ardell, Kuhl, & Sparks, 1985) Since this auditory component by itself provides virtually no information about phonetic structure, it must be the temporal relationships between the auditory and visual channels that contribute to intelligibility (McGrath & Summerfield, 1985). Thus audiovisual speech perception is often more than the sum of its parts; in terms of Massaro's (in press) model, the separate sources are integrated *before* central evaluation. The close integration of inputs from the two modalities is witnessed by anecdotal reports that voicing-triggered buzz accompanying lipreading may assume phonetic qualities (Summerfield, in press).

The theoretical issues raised by audiovisual integration have been discussed thoroughly by Summerfield (in press). He, too, concludes that auditory and visual cues to linguistic structure are integrated before any categorical decisions are made. There are four ways of conceptualizing how this integration occurs: (1) The two channels make independent contributions to linguistic decisions, but temporal relationships provide a third source of information. (2) The visual information is translated into an auditory metric of vocal tract area functions. (3) The auditory information is translated into a visual metric of articulatory kinematics. (4) Both are translated into an abstract representation of dynamic control parameters of articulation. This last-mentioned approach (e.g., Browman & Goldstein, 1986; Kelso, Saltzman, & Tuller, 1986) may ultimately provide the most economic description of speech information in both modalities, and thus may yield the most appropriate vocabulary in which to describe intermodal integration.

### *5. Higher-level Integration*

Human listeners not only integrate auditory and visual information about a speaker's articulations, but they also bring phonotactic, lexical, syntactic, semantic, and pragmatic expectations to bear on their linguistic decisions, provided the auditory and/or visual input is sufficiently ambiguous to give room to effects of such expectations. Some well-known demonstrations of effects in this category are the "phoneme restoration" phenomenon discovered by Warren (1970) and studied more recently by Samuel (1981), in which lexical expectations fill in missing acoustic information, as it were; the lexical bias effect reported by Ganong (1980) and replicated by Fox (1984), which causes a relative shift in the category boundaries on acoustic word-nonword (e.g., DASH-TASH versus DASK-TASK) continua in favor of word percepts; and the "fluent restorations" in rapid shadowing of semantically anomalous passages (Marslen-Wilson, 1985). These phenomena, and a host of related ones often referred to as "top-down" effects, may be considered general forms of cognitive information integration in speech perception. Indeed, Massaro (in press) has argued that the rules by which such higher-level information is integrated with the "bottom-up" information delivered by the senses are the same by which acoustic (and optic) speech cues are integrated. Others argue that top-down influences should be strictly separated from bottom-up processes—that they represent general cognitive functions that operate outside the autonomous speech module (Fodor, 1983; Liberman & Mattingly, 1985). According to this second view, integration of bottom-up cues to phoneme identity is a fundamentally different process from the integration of bottom-up and top-down information. My own view in this matter is that speech perception at every level requires domain-specific knowledge stored in a perceiver's long-term memory. The processes by which this knowledge is brought to bear upon the sensory input are part of our metaphoric representation of brain function and thus are bound to be general (cf. Repp, 1987b). In the absence of a radically different vocabulary in which to characterize the processes within a module (though such a vocabulary will perhaps emerge from the study of articulatory dynamics and coordination), the postulate of a speech module harks back to the "black box" of behaviorism. It is quite likely, of course, that phonetic perception is modular in the sense that integration of phonetic cues precedes, and is not directly influenced by, higher-level factors. This issue can be addressed empirically (see, e.g., Fodor, 1983; Ganong, 1980; Samuel, 1981; Swinney, 1982). My point here is that integration, whether it occurs inside a module or outside it, is conceptually the same thing: a many-to-one mapping. Indeed, Massaro's (e.g., in press) extensive research suggests that the rules of information integration are independent of modularity.

## III. SEGREGATION

The preceding section has illustrated the pervasiveness of integrative processes in speech perception. Much of perceptual and cognitive processing is convergent, with multiple sources of information contributing to single decisions, be they explicit or implicit. Nevertheless, we also need hypothetical mechanisms to prevent all information from converging onto every decision "node." Even though a perceiver's internal criteria for linguistic category membership will automatically reject irrelevant information, information that does not belong is nevertheless often potentially relevant. Thus, in the often-cited cocktail party situation, the voices of several speakers must be kept apart to avoid semantic and phonetic confusions. Various environmental sounds could simulate phonetic events and need to be segregated from the true speech stream. In the speech signal itself, information pertaining to speaker identity, emotion, room acoustics, etc., needs to be distinguished from the phonetic structure, and the overlapping consequences of segmental articulation need to be sorted out. These segregative processes have an important complementary role to play in speech perception: They ensure that integration is restricted to those pieces of information that belong together. Logically, segregation precedes integration, even though functionally they may be just the two sides of one coin. The more physically similar and intertwined the aspects to be segregated are, the more remarkable the segregative process will seem to us.

### A. Temporal and Spatial Segregation

Without any doubt, there are several factors that enable perceivers to distinguish different sound sources or events, regardless of whether they are speech or not. One of these is temporal separation. Sounds occurring a long time apart will usually not be considered as belonging to the same event, although they may come from the same source. In speech, a few seconds are usually enough to segregate phrases or utterances, and a few hundreds of milliseconds of separation usually prevent integration of acoustic cues into a single phonemic decision. One demonstration of this fact may be found in studies of the distinction between single and geminate stop consonants. In a classic experiment, Pickett and Decker (1960) asked English-speaking subjects to distinguish between utterances such as "topic" and "top pick," varying only the duration of the silent /p/ closure. Between 150 and 300 ms were needed to obtain judgments of two /p/s (and two words) rather than just one; the precise duration depended on the overall speaking rate. (See also Obrecht, 1965; Repp, 1978b; 1979a.) If two different stop consonants follow each other, as in the nonsense word /abda/, about 100 ms of silent closure are needed to prevent integration of the two sets of formant transitions into a single stop consonant percept (e.g., Dorman, Raphael, & Liberman, 1979; Repp, 1978b). Dorman et al. (1979) cued the perception of /p/ in "split" solely by inserting a silent interval between an /s/ noise and the syllable "lit" (a percept that may be said to be a pure temporal integration illusion), and subsequently investigated how much silence was needed before subjects reported hearing "s" followed by "lit." This duration turned out to be as long as 600 ms. A subsequent replication (Repp, 1985b) obtained a shorter but still surprisingly long interval of 300-400 ms. To cite a final example, Tillmann, Pompino-Marschall, and Porzig (1984) investigated how much temporal offset of optically and acoustically presented syllables was needed to destroy the audiovisual integration effect discovered by McGurk and MacDonald (1976). It turned out to be 250-300 ms. These various situations have little in common, which explains the different results. The precise duration of the critical interval for segregation surely depends on many factors and does not reflect any general limits of temporal integration. Rather, within the auditory modality it may be related to the closure durations normally encountered in natural speech (see, e.g., Pickett & Decker, 1960; Repp, 1983a).

Temporal asynchrony is a helpful cue in distinguishing speech from other environmental sounds. This was elegantly demonstrated in a series of studies by Darwin (1984; Darwin & Sutherland, 1984), who investigated under what conditions a pure tone added to one of the (pure-tone) harmonics of a synthetic vowel was treated by listeners as part of the vowel spectrum or as a separate nonspeech event. Darwin showed that, when the tone coincided with the vowel, it affected the perceived vowel quality. However, when the onset of the tone preceded that of the vowel or, to a lesser extent, when its offset lagged behind that of the vowel, listeners excluded it from the phonetic information. Similar principles of segregation or "auditory stream formation" have been demonstrated in the perception of nonspeech sounds by Bregman and Pinker (1978).

Another factor that may cause segregation is spatial separation. In real life, the separation of several simultaneous voices or of speech from background noises is often possible because they are perceived as coming from different locations. In the laboratory, presentation over the two channels of earphones has been used to induce segregation. One interesting case in which this form of spatial separation does *not* seem to prevent integration is split-formant or duplex perception, discussed above. Note, however, that in duplex perception one component of the speech signal (the "chirp") *is* segregated and heard as a separate auditory event; the paradox is that this event is still, at the same time, integrated with the speech in the other ear. (See Bregman, 1987.) There are many other instances, however, particularly those in which there is no temporal overlap between the two signals, where spatial separation is sufficient to disrupt perceptual integration. For example, informal observations suggest that, if the artificial "split" created by concatenating "s" and "lit" with some intervening silence is divided between the two ears, so that "s" occurs in one ear and "lit" in the other, this is exactly what listeners report hearing; that is, there is no /p/ percept any more. Similarly, when nasal-consonant-vowel syllables such as /mi/ or /ni/ are divided between the two ears, so that the nasal murmur occurs in one and the vocalic portion containing the formant transitions in the other, listeners have great difficulty identifying the consonant, or in any case do not perform better than if the two components were presented by themselves (Repp, 1987a). Of course, it is always possible to integrate independent sources of information at a cognitive level. These two examples illustrate the role of spatial separation as a segregating factor. Unfortunately, in real life both temporal and spatial separation are often unavailable as segregating agents, and listeners need additional means of sorting out the incoming stream of auditory information.

## B. Spectral Segregation

When irrelevant (speech or nonspeech) sounds are superimposed on speech, listeners have basically two means of segregation at their disposal: Segregation according to local spectral disparity, and according to spectro-temporal (and, in part, speech-specific) criteria of pattern coherence. There are, of course, many sounds in the environment, including those produced by most musical instruments, that are sufficiently different from speech to be perceived immediately as different sources. Local spectral segregation is not always effective, however, and for good reason: First, some nonspeech events (e.g., the pops of bottles or the hisses of steam valves) are spectrally similar to speech sounds and thus are difficult to separate from them locally. Second, and more importantly, speech itself is composed of acoustic segments of diverse spectral composition, and it would be counterproductive if listeners were prone to segregate them, because these segments more often than not map onto the same linguistic unit. Indeed, perceptual segregation of spectrally dissimilar natural speech components can usually be demonstrated only under special conditions, which rarely occur outside the laboratory. Thus, Cole and Scott (1973) rapidly iterated fricative-vowel syllables and found that listeners sometimes reported two streams of events: a train of fricative noises, and a train of vowels, especially when the vocalic formant transitions were removed. A similar phenomenon was obtained with the repeated syllable /ska/ by Diehl, Kluender, and Parker (1985), who then used their findings to explain the different effects of /spa/ or /ska/ stimuli as adaptors (or precursors) in selective adaptation and pairwise contrast paradigms (Sawusch & Jusczyk, 1981; Sawusch & Nusbaum, 1983). The selective adaptation task requires cyclic repetition of a single stimulus, the adaptor, and thus may produce "streaming" of signal components, so that /spa/ is heard as /s/ and /ba/, with the phonological status of the stop consonant altered. Repp (1981) was able to induce listeners through some training to segregate a fricative noise from a following vowel and "hear out" the spectral quality of the noise. Even the individual formants of vowels may segregate under certain conditions. Thomas, Hill, Carrol, and Garcia (1970) and Warren and Warren (1970) observed that it was difficult to perceive the correct temporal order of four rapidly cycling steady-state vowels, and Dorman, Cutting, and Raphael (1975) found that this was because in such artificial sequences individual formants tend to group together and form separate auditory streams. There are anecdotal reports of phoneticians being able to "hear out" individual formants of vowels (e.g., Halle, Hughes, & Radley, 1957; Schubert, 1982), but this ability has remained rare. Still, these various findings underline the fact that spectrally diverse components of the speech signal are *potentially* segregable; fortunately, however, they are perceptually integrated under normal circumstances.

When two different speech streams co-occur, differences in fundamental frequency, intonation pattern, or voice quality may provide cues for separation, in addition to higher-level factors such as syntactic and semantic continuity. Effects of this kind have been found in classical work on selective attention reviewed by Treisman (1969). More recently, Brokx and Nooteboom (1982) obtained a beneficial effect of differences in fundamental frequency and intonation on the identification of meaningless sentences presented against a background of a read story. In the much more artificial situation of two simultaneous steady-state vowels, Scheffers (1983) and Zwicker (1984) found an improvement in recognition performance when a fundamental frequency difference was introduced. Since the magnitude of the difference beyond one semitone did not seem to play a role, the function of $F_0$ differences in this case seems to be to prevent fusion of the two sounds. Similar, though small, effects of $F_0$ on identification scores have also been obtained in dichotic listening studies using synthetic syllables (Halwes, 1969; Repp, 1976a; Tartter & Blumstein, 1981) or vowels (Zwicker, 1984).

The potential of fundamental frequency ($F_0$) and voice quality cues to segregate *successive* portions of speech has also been demonstrated in the laboratory. The mechanisms studied here must be involved in separating different speakers from each other. Several relevant studies have used stimuli in which perception of a stop consonant rested on the duration of a silent closure interval. Dorman et al. (1979) found that when the speech on each side of the silence was produced by different voices, the silence lost its perceptual effectiveness; that is, listeners did not integrate across it. On the other hand, Rakerd, Dechovitz, and Verbrugge (1982) and Verbrugge and Rakerd (1986) have shown that silence retains its effectiveness between syllables produced by male and female voices if the general articulatory and intonational pattern is continuous across the two speakers (achieved by cross-splicing two intact utterances). When the second syllable was spliced onto a first syllable originally produced in utterance-final position, however, the phonetic effect of the silence was disrupted. Thus it seems that dynamic spectro-temporal information about articulatory continuity can override differences in $F_0$ or voice quality. A disruptive effect of discontinuities in intonation on stop consonant perception has also been reported by Price and Levitt (1983), but such an effect was absent in a recent study (Repp, 1985a) in which a constant fricative noise preceded the critical silence, suggesting that the breaks in the $F_0$ contour are effective only when voiced signal portions immediately abut the silent closure interval.

### C. Segregation of Linguistic and Paralinguistic Information

So far I have discussed segregation of two kinds: One separates speech from other, irrelevant sounds (including competing speech streams), and the other dissociates consecutive parts of the same speech stream—a laboratory-induced phenomenon to be avoided in natural speech communication. These segregative processes are "literal" in that they result in the perception of separate sound sources. Segregative processes are also essential, however, when listening to a single speech source, and for two reasons. First, the speech signal conveys in parallel, and largely over the same time-frequency channels, information about phonetic composition, speaker characteristics (vocal tract size, sex, age, identity, emotion), and room or transmission characteristics (reverberation, distortion, filtering). A listener needs to separate these three kinds of information, which Chistovich (1985) has termed "phonetic quality," "personal quality," and "transmission quality," respectively. (See also Traunmüller, 1987.) Second, the acoustic information for adjacent phonemes is overlapped and merged, a phenomenon commonly referred to as coarticulation or "encoding." If phonemic units are to be recovered, the information pertaining to one phoneme needs to be separated from that for another—or so it seems. Both these kinds of segregation are not literal in the sense that they make a speech stream disintegrate perceptually; rather, they separate different aspects of a coherent perceptual event by relating these aspects to different conceptual categories or dimensions represented in long-term memory. They operate on the information in the signal, not on the signal itself.

Of the various types of information segregation of the first kind, that of separating vocal tract size information from phonetic information has received the most attention under the heading of speaker normalization. An explicit solution to this problem is of vital importance to automatic speech recognition as well as to any theory of speech perception. In fact, the focus has been so exclusively on the speaker-independent recovery of phonetic information that it is sometimes forgotten that listeners extract several kinds of information in parallel. Rather than "normalizing" their internal representation of the speech wave and discarding information in the process, they presumably use all available kinds of information to mutual advantage.

Studies of speaker normalization have, for the most part, been concerned with vowels rather than consonants, and with acoustic analysis and automatic recognition rather than with human perception. Older normalization algorithms often required knowledge of a speaker's whole vowel space or average formant frequencies (see Disner, 1980), whereas more recent work has focused on perceptually more relevant transformations based on parameters that are immediately available in the incoming speech signal (e.g., Suomi, 1984; Syrdal & Gopal, 1986; Traunmüller, 1984a). There have been relatively few perceptual studies on this topic; the general assumption has been that it is sufficient to define acoustic properties that are relatively speaker-invariant and also plausible in the light of what is known about the auditory system. Demonstrations of "perceptual normalization" usually show a performance decrement in a listening situation where speaker characteristics are varied rapidly and unpredictably, compared to one in which the speaker remains constant (e.g., Ladefoged & Broadbent, 1957; Summerfield & Haggard, 1975; Verbrugge, Strange, Shankweiler, & Edman, 1976). Although emphasis is sometimes placed on the perceptual "advantage" resulting from effective normalization, the negative consequences of presenting contrived and misleading stimuli are perhaps the more salient outcome of this research (which is by no means unique in this respect).

Analogous experiments have been conducted on normalization in the temporal domain—that is, on the perceptual separation of speaking rate from phonetic length (see review by Miller, 1981). An especially interesting question arises in research on tone languages, where the listener must segregate lexical tones from the overall intonation contour (e.g., Connell, Hogan, & Rozsypal, 1983) and from speaker-dependent variation in $F_0$ (Leather, 1983). In that connection, it is noteworthy that there is mounting evidence (reviewed by Ross, Edmondson, & Seibert, 1986) that tone and intonation perception (and production) are controlled by opposite hemispheres of the brain. At least some forms of linguistic/paralinguistic segregation may thus have a basis in neurophysiological compartmentalization. A general conclusion to be drawn from research on perceptual normalization is that the auditory parameters underlying phonetic classification are not absolute quantities but *relationships* in the spectral and/or temporal domain, computed over a relatively restricted temporal interval, whereas properties signalling speaker sex or identity, emotion, speaking rate, etc., accumulate over longer stretches of speech and/or are based on more nearly absolute quantities.

### D. Segregation of Intertwined Linguistic Information

The emphasis on linguistic information in the vast majority of speech perception studies makes it difficult to find good examples of research on perceptual segregation of linguistic *and* (rather than *from*) nonlinguistic information. Examples of segregation of equivalent information are easier to find when only linguistic information is involved. This leads me to the final topic, one that has been of enormous significance in speech perception research—the problem of *segmentation*, that is, the perceptual separation of the overlapped acoustic correlates of adjacent phonemic units, particularly of vowels and consonants.

One traditional view of the listener's task has been that it is one of phoneme (or feature) extraction, including "compensation" for contextual influences on a segment's acoustic correlates (see the critique by Fowler, 1986). Numerous studies have shown that listeners perceive segments as if they knew all the contextual modifications their acoustic representations undergo (see Repp, 1982; Repp & Liberman, 1987). Thus, for example, a fricative noise ambiguous between /s/ and /ʃ/ in isolation is perceived as /s/ when followed by /u/ but as /ʃ/ when followed by /a/ (Mann & Repp, 1980). One way of describing this finding is that listeners "know" that anticipatory liprounding for /u/ may lower the spectrum of a preceding fricative noise, so they adopt a different criterion for the /s/-/ʃ/ distinction in that context. This view, which emphasizes the role of tacit phonetic knowledge in speech perception, has recently been elaborated by such authors as Flege (in press) and Repp (1987b). The perceptual accomplishment seems more integrative than segregative from that perspective.

An alternative view, having an equally long history, has a recent proponent in Fowler (1984, 1986; Fowler & Smith, 1986) who has likened the separation of overlapping segmental information to mathematical vector analysis. According to her theory, listeners literally subtract or factor out the influences of one segment on another, so that invariant segments are "heard." Fowler conceives of phonetic segments as articulatory events, not as abstract mental categories (see the exchange on coarticulation between Fowler, 1980, 1983, and Hammarberg, 1982), though listeners are assumed to be able to judge their "sound" (Fowler, 1984). Several experiments by Fowler (1981, 1984; Fowler & Smith, 1986) were intended to demonstrate this. They showed that subjects judge acoustically different representations of a segment to be more similar than acoustically identical ones if the former occur in their original contexts while the latter have been spliced into inappropriate contexts. However, since only the former match what listeners expect to hear in a given context, these results are also compatible with an alternative account based on tacit knowledge of contextual effects in speech production (e.g., Repp, 1982; 1987b). That is, rather than having access to the *sound* of segments (Fowler, 1984), listeners may have made their judgments on the basis of the discrepancy of the input from context-sensitive mental norms or prototypes.

Other recent experiments in a similar vein have addressed the separation of nasality and vowel height information in nasalized vowels. Kawasaki (1986) showed that English listeners judge vowels in /m_m/ environment as increasingly nasal as the surrounding nasal murmurs are attenuated; that is, when the nasal consonants are intact, the vowel nasality is attributed to (coarticulation with) the nasal consonants, as it were, and is "factored out" from the vowel percept. Building on this result, Beddor, Krakow, and Goldstein (1986) first established that there are different category boundaries on synthesized /bɛd/-/bæd/ and /bɛ̃d/-/bæ̃d/ continua. English listeners apparently interpret some of the spectral consequences of nasalization as a change in vowel height. However, when an appropriate "conditioning environment" was added in the form of a postvocalic /n/, the category boundary on the resulting /bɛ̃nd/-/bæ̃nd/ continuum was identical with that on the /bɛd/-/bæd/ continuum, as if listeners attributed the vowel nasality to (coarticulation with) the nasal consonant and "factored it out" in Fowler's sense. The result is equally compatible, however, with a theory that postulates context-sensitive vowel (or syllable) prototypes. Indeed, it may be difficult to come up with any decisive experiments. Mentalism and realism may simply represent different metatheoretical perspectives.

Current efforts at Haskins Laboratories to model articulation as a sequence of overlapping segmental gestures (e.g., Browman & Goldstein, 1986; Kelso et al., 1986) may ultimately provide ways of recovering these gestures from the acoustic signal and thus provide a machine implementation of Fowler's vector-analytic concept. A promising mathematical technique for achieving the same goal, based on principal components analysis of vocal tract area function parameters, has been proposed by Atal (1983) and is currently being explored by Marcus (Marcus & Atal, 1986; Marcus & Van Lieshout, 1984). The recovery of articulatory parameters from the acoustic signal remains a central problem in speech research because phonemes and alphabets surely represent an articulatory, not an acoustic classification. However, while a solution of this problem would bring us a great step forward, processes of integration and segregation would still be needed to translate the articulatory "score" into a sequence of discrete segments.

## IV. SPEECH PERCEPTION WITHOUT INTEGRATION AND SEGREGATION?

In the introduction, I discussed four basic assumptions: the separation of the physical and mental worlds, the existence of physical units, the existence of mental units, and the existence of processes relating the two kinds of units. Can a theory of speech perception do without them? The assumptions are not independent, of course: If the physical and mental worlds are distinct, they must receive different descriptions; to be easily communicable in the scientific world, these descriptions must be in terms of discrete concepts or units; and this results in certain functions or relationships between the two descriptive domains. If the physical and mental worlds were isomorphic, there would be no need for a theory of perception. If one or the other description were without units (more likely an error of omission than a deliberate theoretical choice), then perception would seem either entirely integrative or entirely segregative—not an attractive state of affairs. Denial of functions, however abstract, linking the two domains would merely impoverish perceptual theory. Certainly we need these functions in theories of auditory processing and organization. As to the perception of phonetic information, however, an alternative approach has been proposed.

This approach, stated most eloquently by Studdert-Kennedy (1985) and Fowler (1986), follows the "direct-realist" perspective of ecological psychology (see, e.g., Gibson, 1979; Warren & Shaw, 1985). Although it affirms the existence of linguistic units as articulatory events, it essentially abandons the distinction between the physical and mental domains. The segmental structure of speech (as characterized by the linguist or phonetician) is assumed to be ever-present on its way from the speaker's to the listener's brain. There is assumed to be a direct isomorphism between physical and mental descriptions of speech events (such as phonemes), though it is acknowledged that the appropriate physical and motor-dynamic descriptions have not been fully worked out. Thus this school of thought rejects the idea that the input is divided into parts that need to be integrated or segregated by the listener; rather, the input units are taken to be identical with the perceptual units—that is, they *are already* integrated or segregated with respect to more primitive acoustic or auditory units. The deliberate strategy of this philosophy is to eliminate classical problems in perceptual research (such as segmentation and invariance) by redefining and redescribing physical events. Rather than being attributed to the perceiver's brain, the burdens of information integration and segregation thus fall upon the investigator trying to find an "integral" description of "separate" speech events. However, this effort is equivalent to that of finding a principled explanation of perceptual integration and segregation: If we can show that certain pieces of input are always integrated, we might as well call them integral and treat them as a single piece in our descriptions—if we only had names for them. Behind the rhetoric and the different terminologies of mentalistic and realistic approaches lies a common goal: to arrive at the most economic characterization of linguistic structure in all its physical incarnations. Clearly, even speech research propelled by a mentalistic philosophy (still predominant in the field) must strive to minimize the work attributed to a speaker-listener's mind. But will we be able to relieve it of its entire burden to integrate and segregate? What we take away (in theory) is likely to re-emerge as logical conjunctions, disjunctions, and relational terms in our physical characterization of speech events. As long as we scientists communicate in conventional language, integration and segregation at some stage in our theories will be difficult to avoid.

## References

Assmann, P. F., & Nearey, T. M. (1987). Perception of front vowels: The role of harmonics in the first formant region. *Journal of the Acoustical Society of America, 81*, 520-534.

Atal, B. S. (1983). Efficient coding of LPC parameters by temporal decomposition. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (Boston)*, 81-84.

Bailey, P. J., & Summerfield, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 536-563.

Beddor, P. S. (1984). Formant integration and the perception of nasal vowel height. *Haskins Laboratories Status Report on Speech Research, SR-77/78*, 107-120.

Beddor, P. S., Krakow, R. A., & Goldstein, L. M. (1986). Perceptual constraints and phonological change: A study of nasal vowel height. *Phonology Yearbook, 3,* 197-218.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics, 29,* 191-211.

Bladon, A. (1982). Arguments against formants in the auditory representation of speech. In R. Carlson & B. Granström (Eds.), *The representation of speech in the peripheral auditory system* (pp. 95-102). Amsterdam: Elsevier Biomedical Press.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America, 67,* 648-662.

Breeuwer, M., & Plomp, R. (1986). Speechreading supplemented with auditorily presented speech parameters. *Journal of the Acoustical Society of America, 79,* 481-499.

Bregman, A. S. (1978). The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII* (pp. 63-76). Hillsdale, NJ: Erlbaum.

Bregman, A. S. (1987). The meaning of duplex perception: Sounds as transparent objects. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception*. The Hague: Martinus Nijhoff.

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology, 32,* 19-31.

Bregman, A. S., & Steiger, H. (1980). Auditory streaming and vertical localization: Interdependence of "what" and "where" decisions in audition. *Perception & Psychophysics, 28,* 539-546.

Broadbent, D. E., & Ladefoged, P. (1957). On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America, 29,* 708-710.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3,* 219-254.

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics, 10,* 23-36.

Bryan, W. L., & Harter, N. (1899). Studies in the physiology and psychology of the telegraphic language. The acquisition of a hierarchy of habits. *Psychological Review, 6,* 345-375.

Carlson, R., Fant, G., & Granström, B. (1975). Two-formant models, pitch, and vowel perception. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 55-82). London: Academic Press.

Carlson, R., Granström, B., & Fant, G. (1970). Some studies concerning perception of isolated vowels. *Speech Transmission Laboratory Quarterly Progress and Status Report, 2-3,* 19-35 (Stockholm: Royal Technical University).

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62,* 961-970.

Celmer, R. D., & Bienvenue, G. (1987). Critical bands in the perception of speech by normal and sensorineural hearing loss listeners. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception*. The Hague: Martinus Nijhoff.

Chistovich, L. A. (1985). Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America, 77,* 789-805.

Chistovich, L. A., & Lublinskaja, V. V. (1979). The center of gravity effect in vowel spectra and critical distance between the formants. *Hearing Research, 1,* 185-195.

Cole, R. A., & Scott, B. (1973). Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology, 27,* 441-449.

Connell, B. A., Hogan, J. T., & Rozsypal, A. J. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics, 11,* 337-351.

Cowan, N. (in press). Auditory sensory storage in relation to the growth of sensation and acoustic information extraction. *Journal of Experimental Psychology: Human Perception and Performance.*

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review, 83,* 114-140.

Darwin, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset-time. *Quarterly Journal of Experimental Psychology, 33A,* 185-207.

Darwin, C. J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of the Acoustical Society of America, 76,* 1636-1647.

Darwin, C. J., & Gardner, R. B. (1985). Which harmonics contribute to the estimation of first formant frequency? *Speech Communication, 4,* 231-235.

Darwin, C. J., & Seton, J. (1983). Perceptual cues to the onset of voiced excitation in aspirated initial stops. *Journal of the Acoustical Society of America, 74,* 1126-1135.

Darwin, C. J., & Sutherland, N. S. (1984). Grouping frequency components of vowels: When is a harmonic not a harmonic? *The Quarterly Journal of Experimental Psychology, 36A,* 193-208.

Delattre, P., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word, 8,* 195-210.

Delgutte, B. (1980). Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *Journal of the Acoustical Society of America, 68,* 843-857.

Delgutte, B., & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America, 75,* 897-907.

Derr, M. A., & Massaro, D. W. (1980). The contribution of vowel duration, Fo contour, and frication duration as cues to the /juz/-/jus/ distinction. *Perception & Psychophysics, 27,* 51-59.

Deutsch, D. (1978). Lateralization by frequency for repeating sequences of dichotic 400- and 800-Hz tones. *Journal of the Acoustical Society of America, 63,* 184-186.

Deutsch, D., & Roll, P. (1976). Separate "what" and "where" decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 23-29.

Diehl, R. (1987). Auditory constraints on the perception of speech. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception.* The Hague: Martinus Nijhoff.

Diehl, R. L., Kluender, K. R., & Parker, E. M. (1985). Are selective adaptation and contrast effects really distinct? *Journal of Experimental Psychology: Human Perception and Performance, 11,* 209-220.

Disner, S. F. (1980). Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America, 67,* 253-261.

Dorman, M. F., Cutting, J. E., & Raphael, L. J. (1975). Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 121-129.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. (1979). Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America, 65,* 1518-1532.

Dreschler, W. A., & Plomp, R. (1980). Relation between psychophysical data and speech perception for hearing-impaired subjects. I. *Journal of the Acoustical Society of America, 68,* 1608-1615.

Eggermont, J. J. (1985). Peripheral auditory adaptation and fatigue: A model oriented review. *Hearing Research, 18,* 57-71.

Elman, J. L., & McClelland, J. L. (1984). Speech perception as a cognitive process: The interactive activation model. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice. Vol. 10* (pp. 337-373). New York: Academic Press.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics, 27,* 343-350.

Flege, J. E. (in press). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders, Vol. 1.* Norwood, NJ: Ablex.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8,* 113-133.

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research, 46,* 127-139.

Fowler, C. A. (1983). Realism and unrealism: a reply. *Journal of Phonetics, 11,* 303-322.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics, 36,* 359-368.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14,* 3-28.

Fowler, C. A., & Smith, M. R. (1986). Speech perception as "vector analysis": An approach to the problems of invariance and segmentation. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123-135). Hillsdale, NJ: Erlbaum.

Fox, R. A. (1983). Perceptual structure of monophthongs and diphthongs in English. *Language and Speech, 26,* 21-60.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 526-540.

Ganong, W. F., III. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110-125.

Garner, W. R. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Grant, K. W., Ardell, L. H., Kuhl, P. K., & Sparks, D. W. (1985). The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects. *Journal of the Acoustical Society of America, 77,* 671-677.

Green, D. M. (1983). Profile analysis. A different view of auditory intensity discrimination. *American Psychologist, 38,* 133-142.

Green, D. M., & Kidd, G., Jr. (1983). Further studies of auditory profile analysis. *Journal of the Acoustical Society of America, 73,* 1260-1265.

Green, K. P., & Miller, J. L. (1985). On the role of visual rate information in phonetic perception. *Perception & Psychophysics, 38,* 269-276.

Halle, M., Hughes, G. W., & Radley, J.-P. A. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America, 29,* 107-116.

Halwes, T. G. (1969). Effects of dichotic fusion on the perception of speech. *Haskins Laboratories Status Report on Speech Research (Supplement).*

Hammarberg, R. (1982). On redefining coarticulation. *Journal of Phonetics, 10,* 123-137.

Jaeger, J. J. (1980). Testing the psychological reality of phonemes. *Language and Speech, 23,* 233-253.

Jenkins, J. J. (1985). Acoustic information for objects, places, and events. In W. H. Warren & R. E. Shaw (Eds.), *Persistence and change. Proceedings of the First International Conference on Event Perception* (pp. 115-138). Hillsdale, NJ: Erlbaum.

Jongman, A., Blumstein, S. E., & Lahiri, A. (1985). Acoustic properties for dental and alveolar stop consonants: a cross-language study. *Journal of Phonetics, 13,* 235-251.

Kawasaki, H. (1986). Phonetic explanation of phonological universals: The case of distinctive vowel nasalization. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 81-104). New York: Academic Press.

Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics, 14,* 29-59.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America, 73,* 322-335.

Kewley-Port, D., Pisoni, D. B., & Studdert-Kennedy, M. (1983). Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America, 73,* 1779-1793.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7,* 279-312.

Kohler, K. J. (1979). Dimensions in the perception of fortis and lenis plosives. *Phonetica, 36,* 332-343.

Kubovy, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 55-98). Hillsdale, NJ: Erlbaum.

Kunisaki, O., & Fujisaki, H. (1977). On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics, 11,* 85-91 (University of Tokyo).

Kurowski, K., & Blumstein, S. E. (in press). Acoustic properties for place of articulation in nasal consonants. *Journal of the Acoustical Society of America.*

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29,* 98-104.

Lahiri, A., Gewirth, L., & Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study. *Journal of the Acoustical Society of America, 76,* 391-404.

Leather, J. (1983). Speaker normalization in perception of lexical tone. *Journal of Phonetics, 11,* 373-382.

Lehiste, I. (1972). The units of speech perception. In J. H. Gilbert (Ed.), *Speech and cortical functioning* (pp. 187-236). New York: Academic Press.

Liberman, A. M. (1979). Duplex perception and integration of cues: Evidence that speech is different from nonspeech and similar to language. In E. Fischer-Jorgensen, J. Rischel, &

N. Thorsen (Eds.), *Proceedings of the Ninth International Congress of Phonetic Sciences* (pp. 468-473). Copenhagen: Institute of Phonetics, University of Copenhagen.

Liberman, A. M. (1982). On finding that speech is special. *American Psychologist, 37,* 148-167.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language, 33,* 42-49.

Lisker, L. (1978). On buzzing the English /b/. *Haskins Laboratories Status Report on Speech Research, SR-55/56*, 181-188.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24,* 253-257.

Mann, V. A. (1986). Phonological awareness: The role of reading experience. *Cognition, 24,* 65-92.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics, 28,* 213-228.

Marcus, S. M., & Atal, B. S. (1986). Decoding the speech code—applications of temporal decomposition. *Journal of the Acoustical Society of America, 80* (Suppl. 1), S17.

Marcus, S. M., & Van Lieshout, R. A. J. M. (1984). Temporal decomposition of speech. *IPO Annual Progress Report*, 25-31 (Nijmegen, The Netherlands).

Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication, 4,* 55-73.

Massaro, D. W. (1975). Preperceptual images, processing time, and perceptual units in speech perception. In D. W. Massaro (Ed.), *Understanding language. An information-processing analysis of speech perception, reading, and psycholinguistics* (pp. 125-150). New York: Academic Press.

Massaro, D. W. (in press). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Erlbaum.

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 753-771.

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America, 67,* 996-1013.

Mattingly, I. G. (1972). Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye. The relationships between speech and reading* (pp. 133-148). Cambridge, MA: MIT Press.

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America, 77,* 678-685.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

McNeill, D., & Lindig, K. (1973). The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior, 12,* 419-430.

Miller, J. L. (1981). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech.* Hillsdale, NJ: Erlbaum.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25,* 457-465.

Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidth and excitation patterns. *Journal of the Acoustical Society of America, 74,* 750-753.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7,* 323-331.

Mushnikov, V. N., & Chistovich, L. A. (1973). Experimental testing of the band hypothesis of vowel perception. *Soviet Physics-Acoustics, 19,* 250-254.

Obrecht, D. H. (1965). Three experiments in the perception of geminate consonants in Arabic. *Language and Speech, 8,* 31-41.

Parker, E. M., Diehl, R. L., & Kluender, K. R. (1986). Trading relations in speech and nonspeech. *Perception & Psychophysics, 39,* 129-142.

Peters, R. W., Moore, B. C. J., & Glasberg, B. R. (1983). Pitch of components of complex tones. *Journal of the Acoustical Society of America, 73,* 924-929.

Pickett, J. M., & Decker, L. R. (1960). Time factors in perception of a double consonant. *Language and Speech, 3,* 11-17.

Pisoni, D. B. (1987). Auditory perception of complex sounds: Some comparisons of speech vs. non-speech signals. In W. A. Yost & C. S. Watson (Eds.), *Auditory processing of complex sounds* (pp. 247-256). Hillsdale, NJ: Erlbaum.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception & Psychophysics, 34,* 314-322.

Plomp, R. (1964). The ear as a frequency analyzer. *Journal of the Acoustical Society of America, 36,* 1628-1636.

Pols, L. C. W., van der Kamp, L. J. Th., & Plomp, R. (1969). Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America, 46,* 458-467.

Port, R. F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics, 7,* 45-56.

Price, P. J., & Levitt, A. G. (1983). The relative roles of syntax and prosody in the perception of the /š/-/č/ distinction. *Language and Speech, 26*, 291-304.

Rakerd, B., Dechovitz, D. R., & Verbrugge, R. R. (1982). An effect of sentence finality on the phonetic significance of silence. *Language and Speech, 25,* 267-282.

Rakerd, B., & Verbrugge, R. R. (1985). Linguistic and acoustic correlates of the perceptual structure found in an individual differences scaling study of vowels. *Journal of the Acoustical Society of America, 77,* 296-301.

Rand, T. C. (1974). Dichotic release from masking for speech. *Journal of the Acoustical Society of America, 55,* 678-680.

Repp, B. H. (1976a). Effects of fundamental frequency contrast on discrimination and identification of dichotic CV syllables at various temporal delays. *Memory & Cognition, 4,* 75-90.

Repp, B. H. (1976b). Identification of dichotic fusions. *Journal of the Acoustical Society of America, 60,* 456-469.

Repp, B. H. (1978a). *"Coperception": Influence of vocalic context on same-different judgments about intervocalic stop consonants.* Unpublished manuscript (available from the author).

Repp, B. H. (1978b). Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics, 24,* 471-485.

Repp, B. H. (1979a). Influence of vocalic environment on perception of silence in speech. *Haskins Laboratories Status Report on Speech Research, SR-57,* 267-290.

Repp, B. H. (1979b). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech, 22,* 173-189.

Repp, B. H. (1980). Accessing phonetic information during perceptual integration of temporally distributed cues. *Journal of Phonetics, 8,* 185-194.

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception & Psychophysics, 30,* 217-227.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92,* 81-110.

Repp, B. H. (1983a). Bidirectional contrast effects in the perception of VC-CV sequences. *Perception & Psychophysics, 33,* 147-155.

Repp, B. H. (1983b). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication, 2,* 341-362.

Repp, B. H. (1984a). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice. Vol. 10* (pp. 243-335). New York: Academic Press.

Repp, B. H. (1984b). Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. *Language and Speech, 27,* 245-254.

Repp, B. H. (1984c). The role of release bursts in the perception of [s]-stop clusters. *Journal of the Acoustical Society of America, 75,* 1219-1230.

Repp, B. H. (1985a). Can linguistic boundaries change the effectiveness of silence as a phonetic cue? *Journal of Phonetics, 13,* 421-431.

Repp, B. H. (1985b). Perceptual coherence of speech: Stability of silence-cued stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 799-813.

Repp, B. H. (1987a). *On the possible role of auditory short-term adaptation in perception of the prevocalic [m]-[n] contrast.* Manuscript submitted for publication.

Repp, B. H. (1987b). The role of psychophysics in understanding speech perception. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception.* The Hague: Martinus Nijhoff.

Repp, B. H., & Bentin, S. (1984). Parameters of spectral/temporal fusion in speech perception. *Perception & Psychophysics, 36,* 523-530.

Repp, B. H., & Liberman, A. M. (1987). Phonetic category boundaries are flexible. In S. N. Harnad (Ed.), *Categorical perception.* New York: Cambridge University Press.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 621-637.

Repp, B. H., Milburn, C., & Ashkenas, J. (1983). Duplex perception: Confirmation of fusion. *Perception & Psychophysics, 33,* 333-337.

Rosen, S. M., Fourcin, A. J., & Moore, B. C. J. (1981). Voice pitch as an aid to lipreading. *Nature, 291,* 150-152.

Ross, E. D., Edmondson, J. A., & Seibert, G. B. (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice. *Journal of Phonetics, 14,* 283-302.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110,* 474-494.

Savin, H. B., & Bever, T. G. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior, 9,* 295-302.

Sawusch, J. R., & Jusczyk, P. (1981). Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human Perception and Performance,* 7, 408-421.

Sawusch, J. R., & Nusbaum, H. C. (1983). Auditory and phonetic processes in place perception for stops. *Perception & Psychophysics, 34,* 560-568.

Scheffers, M. T. M. (1983). *Sifting vowels. Auditory pitch analysis and sound segregation.* Unpublished doctoral dissertation, University of Groningen, The Netherlands.

Schubert, E. D. (1982). On hearing your own performance. In V. L. Lawrence (Ed.), *Transcripts of the Eleventh Symposium Care of the Professional Voice* (pp. 161-185). New York: The Voice Foundation.

Schwartz, J. L., & Escudier, P. (1987). Does the human auditory system include large scale spectral integration? In M. E. H. Schouten (Ed.), *The psychophysics of speech perception.* The Hague: Martinus Nijhoff.

Stelmachowicz, P. G., Jesteadt, W., Gorga, M. P., & Mott, J. (1985). Speech perception ability and psychophysical tuning curves in hearing-impaired listeners. *Journal of the Acoustical Society of America, 77,* 620-627.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 64,* 1358-1368.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (eds.), *Perspectives in the study of speech* (pp. 1-38). Hillsdale, NJ: Erlbaum.

Studdert-Kennedy, M. (1985). Perceiving phonetic events. In W. H. Warren, Jr., & R. E. Shaw (Eds.), *Persistence and change: Proceedings of the First International Conference on Event Perception* (pp. 139-156). Hillsdale, NJ: Erlbaum.

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica, 36,* 314-331.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1074-1095.

Summerfield, Q. (1983). Audio-visual speech perception, lipreading, and artificial stimulation. In M. E. Lutman & M. P. Haggard (Eds.), *Hearing science and hearing disorders* (pp. 131-182). London: Academic Press.

Summerfield, Q. (in press). Preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye.* Hillsdale, NJ: Erlbaum.

Summerfield, Q., & Assmann, P. (1987). Auditory enhancement and speech perception. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception.* The Hague: Martinus Nijhoff.

Summerfield, Q., & Haggard, M. P. (1975). Vocal tract normalization as demonstrated by reaction times. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 115-142). London: Academic Press.

Summerfield, Q., Haggard, M., Foster, J., & Gray, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. *Perception & Psychophysics, 35,* 203-213.

Suomi, K. (1984). On talker and phoneme information conveyed by vowels: A whole spectrum approach to the normalization problem. *Speech Communication*, 3, 199-209.

Suomi, K. (1985). The vowel-dependence of gross spectral cues to place of articulation of stop consonants in CV syllables. *Journal of Phonetics, 13,* 267-285.

Swinney, D. A. (1982). The structure and time-course of information interaction during speech comprehension: Lexical segmentation, access, and interpretation. In J. Mehler, E. C. T. Walker, & M. Garrett (Eds.), *Perspectives on mental representation. Experimental and* *theoretical studies of cognitive processes and capacities* (pp. 151-167). Hillsdale, NJ: Erlbaum.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79,* 1086-1100.

Tartter, V. C., & Blumstein, S. E. (1981). The effects of pitch and spectral differences on phonetic fusion in dichotic listening. *Journal of Phonetics, 9,* 251-259.

Tartter, V. C., Kat, D., Samuel, A. G., & Repp, B. H. (1983). Perception of intervocalic stop consonants: The contributions of closure duration and formant transitions. *Journal of the Acoustical Society of America, 74,* 715-725.

Thomas, I. B., Hill, P. B., Carrol, F. S., & Garcia, D. (1970). Temporal order in the perception of vowels. *Journal of the Acoustical Society of America, 48,* 1010-1013.

Tillmann, H. G., Pompino-Marschall, B., & Porzig, U. (1984). Zum Einfluss visuell dargebotener Sprechbewegungen auf die Wahrnehmung der akustisch kodierten Artikulation. *Forschungsbericht des Instituts für Phonetik und Sprachliche Kommunikation der Universität München, 19,* 318-336.

Tomiak, G. R., Mullennix, J. W., & Sawusch, J. R. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America, 81,* 755-764.

Traunmüller, H. (1982). Perception of timbre: Evidence for spectral resolution bandwidth different from critical band? In R. Carlson & B. Granström (Eds.), *The representation of speech in the peripheral auditory system* (pp. 103-108). Amsterdam: Elsevier Biomedical Press.

Traunmüller, H. (1984a). Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels. *Speech Communication*, 3, 49-61.

Traunmüller, H. (1984b). Die spektrale Auflösung bei der Wahrnehmung der Klangfarbe von Vokalen. *Acustica, 54,* 237-246.

Traunmüller, H. (1987). Some aspects of the sound of speech sounds. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception*. The Hague: Martinus Nijhoff.

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review, 76,* 282-299.

Verbrugge, R. R., & Rakerd, B. (1986). Evidence for talker-independent information for vowels. *Language and Speech, 29,* 39-57.

Verbrugge, R. R., Strange, W., Shankweiler, D. P., & Edman, T. R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America, 60,* 198-212.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167,* 392-393.

Warren, R. M., & Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American, 233,* 30-36.

Warren, W. H., Jr., & Shaw, R. E. (1985). *Persistence and change: Proceedings of the First International Conference on Event Perception*. Hillsdale, NJ: Erlbaum.

Warren, W. H., Jr., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 704-712.

Watson, C. S., & Foyle, D. C. (1985). Central factors in the discrimination and identification of complex sounds. *Journal of the Acoustical Society of America, 78,* 375-380.

Weintraub, M. (1987). Sound separation and auditory perceptual organization. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception*. The Hague: Martinus Nijhoff.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. *Journal of the Acoustical Society of America, 69,* 275-282.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics, 35,* 49-64.

Whalen, D. H., & Liberman, A. M. (1987). *Speech perception takes precedence over nonspeech perception*. Manuscript submitted for publication.

Whalen, D. H., & Samuel, A. G. (1985). Phonetic information is integrated across intervening nonlinguistic sounds. *Perception & Psychophysics, 37,* 579-587.

Wright, H. N. (1964). Temporal summation and backward masking. *Journal of the Acoustical Society of America, 36,* 927-932.

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America, 68,* 1523-1525.

Zwicker, U. T. (1984). Auditory recognition of diotic and dichotic vowel pairs. *Speech Communication, 3,* 265-277.

Zwislocki, J. J. (1969). Temporal summation of loudness: An analysis. *Journal of the Acoustical Society of America, 46,* 431-440.

# SPEECH PERCEPTION TAKES PRECEDENCE OVER NONSPEECH PERCEPTION*

**D. H. Whalen and Alvin M. Liberman†**

**Abstract.** *When made more intense, some components of a speech signal can be heard simultaneously as speech and nonspeech—a form of "duplex" perception—though at lower intensities, the speech alone is heard. Such intensity-dependent duplexity implies the existence of a phonetic mode of perception that takes precedence over auditory modes.*

## INTRODUCTION

One theory of speech perception holds that there is a biologically distinct system, or "module," specialized for extracting phonetic elements—notably, consonants and vowels—from the sounds that convey them (Liberman & Mattingly, 1985). The percepts produced by this module are immediately phonetic in character; accordingly, they stand apart from auditory percepts that are composed of such dimensions as pitch, loudness, and timbre. There is, then, no first-stage auditory percept, as most other theories of speech require (Cole & Scott, 1974; Oden & Massaro, 1978; Stevens, 1975), hence no need for a subsequent stage in which the auditory tokens are matched to phonetic prototypes, and so made appropriate for further processing as language. Indeed, as the experiments reported here show, it is the phonetic module that has priority, as if its processes occurred before, not after, those that yield the standard dimensions of auditory perception.

Consistent with the existence of a distinct phonetic mode is the fact that a particular piece of sound can evoke radically different percepts, depending on whether or not it engages the phonetic module. Consider, for example, acoustic patterns sufficient for synthesizing on a computer the syllables "da" and "ga," as shown at the top of Figure 1. The three formants represent resonances of the vocal tract and have, at their onsets, frequency sweeps called transitions. These transitions last approximately 50 ms and reflect the way the resonances change as the tongue and jaw move from the consonant to the vowel. Normally, the perceived distinction between "da" and "ga" depends on many acoustic variables; as seen in the figure, however, it can be made to depend only on differences in the transition of the third formant. Thus, in the context of the syllable,

---

* *Science*, 1987, *237*, 169-171.

† Also Yale University and University of Connecticut.

*Acknowledgment*. This research was supported by NICHD grant HD-01994 to Haskins Laboratories. We thank Carol A. Fowler, Ignatius Mattingly, Bruno Repp, Lawrence Rosenblum, Philip Rubin, and Michael Studdert-Kennedy for helpful comments.

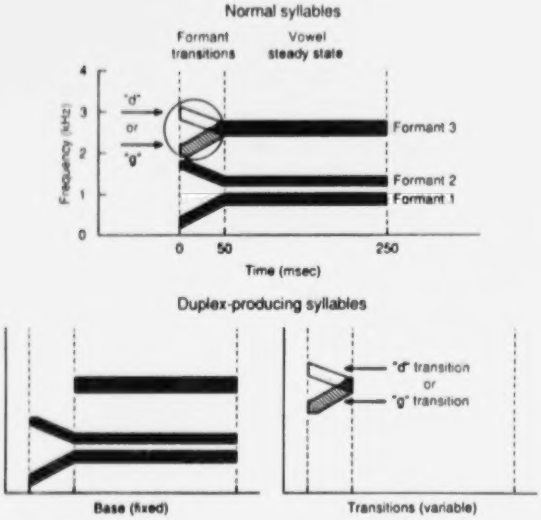

**Figure 1.** Schematic representation of the syllables.

these transitions become crucial to the phonetic percept. But in isolation (as at bottom right of the figure) they are heard as the glissandi or differently pitched "chirps" that psychoacoustic considerations would lead one to expect. These two ways of perceiving the formant transitions—one phonetic, the other auditory—are strikingly different: There is no hint of chirpiness in the "da" or "ga," and no da-ness or ga-ness in the chirps; moreover, the transitions are discriminated differently depending on the mode in which they are perceived (Mattingly, Liberman, Syrdal, & Halwes, 1971).

Under special circumstances, the transitions can evoke the phonetic and auditory percepts simultaneously. This curious effect, called "duplex perception," occurs when the third-formant transition is presented by itself to one ear, while the remainder of the pattern, called the "base," (see the bottom left of the figure) is presented to the other. Listeners then simultaneously hear a chirp (in the ear to which the transition is presented) and (in the other ear) the syllable "da" or "ga," as determined by the transition. These simultaneous percepts, and the very different discrimination functions they yield, are very nearly the same as those produced, separately, by the isolated transitions and the whole syllable (Mann & Liberman, 1983).

Since duplex perception occurs in response to a fixed acoustic pattern and results in two simultaneous percepts, it can hardly be attributed to auditory interactions arising from changes in acoustic context or to a shifting of attention between two forms of an ambiguous stimulus. And the fact that the "da" or "ga" is perceived to be entirely in one ear, though the critical transition had been presented only to the other, argues that the incorporation of the transition into the base is an integration at the perceptual level, not a "cognitive" afterthought that deliberately combines what had initially been perceived as separate.

Thus, duplex perception provides support for the view that there are distinct phonetic and auditory ways of perceiving the same (speech) signal, but in so doing, it poses a question that might otherwise have gone unasked: Why, in the normal case, are the components of speech not perceived duplexly—that is, why is the "da" or "ga" not normally accompanied by the chirp?

Relying on considerations of plausibility and parsimony, Mattingly and Liberman (in press) proposed that the phonetic module "preempts" the phonetically relevant parts of the signal before making the remainder available to auditory processing. This proposal seemed plausible, because, in contrast to the indefinitely large set of acoustic events that occur, phonetic events form a natural class that is defined by its correspondence to the acoustic results of specialized movements of the articulatory organs. The proposal was parsimonious because the very processes of phonetic perception remove from the signal all evidence of those phonetic events, and thus preclude such (parallel) processing as would cause them to be perceived yet again as chirps. This "reemptiveness" is similar to the precedence we have spoken of, and that we mean to demonstrate directly with a new and somewhat simpler version of duplex perception. (See Darwin & Sutherland, 1984, p. 206, for a related observation.)

The new procedure differs from the old in that the two parts of the signal are not divided between the ears, but are, rather, presented equally to both. Now duplexity is produced (in both ears at once) by changing the intensity of the transition relative to the base. At relatively low intensities, the transitions serve only their expected phonetic function. At higher intensities, however, the transitions continue to make their phonetic contribution but simultaneously evoke nonspeech "chirps." These observations, which we made initially in pilot experiments, suggested that we test the following generalizations:

1) In isolation, neither transition sounds like "da" or "ga."

2) In syllabic context, the transitions will, at some intensity, evoke nonspeech chirps, establishing a "duplexity threshold."

3) Above the duplexity threshold, the chirps can be matched to those evoked by the transitions in isolation.

4) Both below the duplexity threshold and above it, the transitions appropriately determine whether the syllable is heard as "da" or "ga."

The stimuli were the same as those represented in the figure, except that the third-formant transitions were not frequency bands excited by a fundamental (as were the formants of the base), but, rather, time-varying sinusoids that follow the center frequencies. We had found that such sinusoidal transitions combine with the formant-synthesized base to make coherent phonetic percepts, in this case "da" and "ga." But the sinusoids have the advantage, for our purposes, that in isolation they produce "whistles," which we found to be more easily discriminated than the chirps, and even less speech-like.

The base syllable was created with a software formant synthesizer; the sinusoids were created with another software synthesizer designed for pure-tone generation. From a set of input parameter values representing frequencies and amplitudes, each synthesizer calculated a digital waveform that was then turned into sound via a digital-to-analog converter.

The base was synthesized in one computer file and the two sinusoidal transitions (one modeled after "d" and one after "g") in two other files. The base and one transition could then be output through synchronized D-to-A channels, separately attenuated, and electronically combined for presentation over headphones as a single sound to subjects. The base was presented at a fixed intensity of 72 dB SPL.

Eleven young adult speakers of English (six female and five male) with no reported hearing problems were run in separate sessions. None knew anything about the composition of the stimuli or the purpose of the experiment. They were paid for their participation. One failed to perceive in a duplex fashion at the intensity levels available, and so was excluded from all analyses.

Initially, subjects were asked to identify the sinusoidal transitions as "da" or "ga." Twenty repetitions of each were presented in random order. The subjects implied that they considered the request absurd, since, as they insisted, the whistles did not sound at all like speech. They nevertheless complied, with results that are shown in the first column of Table 1. (For all tests, there was no significant difference between the responses to the "d" and "g" stimuli, so only the combined percentages are reported.) Most subjects picked one whistle or the other as "da" and held to that consistently. Some happened to pick the correct one; others were just as consistently wrong. One (S9) simply called all the whistles "da." Overall, identification accuracy did not differ significantly from chance, $t(9) = 1.22$, n.s.

**Table 1**

**Percent correct performance on the four main tasks (results from 40 trials per subject).**

| Subject | Identification of isolated sinusoids as "d" or "g" | Match of "duplex" to isolated sinusoids | Identification of syllables as "da" or "ga" below duplexity threshold | Identification of syllables as "da" or "ga" above duplexity threshold |
|---|---|---|---|---|
| 1 | 72.5 | 92.5 | 100.0 | 100.0 |
| 2 | 100.0 | 65.0 | 100.0 | 97.5 |
| 3 | 15.0 | 97.5 | 100.0 | 100.0 |
| 4 | 95.0 | 97.5 | 100.0 | 100.0 |
| 5 | 30.0 | 85.0 | 97.5 | 100.0 |
| 6 | 95.0 | 72.5 | 92.5 | 85.0 |
| 7 | 100.0 | 87.5 | 82.5 | 97.5 |
| 8 | 0.0 | 95.0 | 52.5 | 100.0 |
| 9 | 50.0 | 47.5 | 100.0 | 97.5 |
| 10 | 90.0 | 65.0 | 100.0 | 100.0 |
| Mean | 64.8 | 80.5 | 92.5 | 97.8 |
| S.E.M. | ±12.1 | ±5.4 | ±4.8 | ±1.5 |

To find the intensity at which the sinusoids in syllabic context evoked nonspeech whistles in addition to "da" or "ga" (the "duplexity threshold"), we had the subjects adjust the attenuator that controlled the intensity of the sinusoid until the whistle was just audible. This was done three times for each sinusoid. The mean duplexity thresholds for all subjects, expressed in relation to the steady-state of the third formant, were -6.4 db (s.d. 5.0 db) for the "da" sinusoid and 0.0 db (s.d. 4.9 db) for the "ga" sinusoid. This difference in duplexity thresholds, which was found for all ten subjects, is consistent with the fact that, in isolation, the "da" sinusoid—the one with the lower duplexity threshold—was louder.

To make sure that the whistle component of the duplex percept was comparable to the whistle of the sinusoid in isolation, we carried out a matching test. On each trial, three stimuli were presented: first, one sinusoid in isolation, then either of the two sinusoids in syllabic context, and finally the other sinusoid in isolation. Each sinusoid occurred with the syllable twenty times, matching the first sinusoid or the last an equal number of times. The sinusoid in the syllable was presented at 6 db above the duplexity threshold for "ga." Subjects judged whether the duplexly perceived whistle was more like the isolated whistle that preceded or followed. As the second column of Table 1 makes clear, subjects were able to do this rather demanding task well above chance, $t(9) = 5.50, p < .001$.[1]

To test whether the sinusoids reliably determined how the syllable was perceived below the duplexity threshold, we set them 4 db below the "da" duplexity threshold and presented twenty repetitions of each in random order. Subjects were to identify the consonant as "d" or "g." Again, they performed well above chance, $t(9) = 8.88, p < .001$, as seen in Table 1, column 3.

It remained, then, to determine that the sinusoids continue to provide phonetic information even when they also evoke whistles. For that purpose, we set the sinusoids at 6 db above the higher ("ga") duplexity threshold and carried out an identification test like the one just described. Comparing the rightmost columns of the table, we see that subjects were no less accurate above the duplexity threshold than below it, $t(9) = 32.60, p < .001$ for Column 4.

Thus, at lower levels of intensity, the sinusoids provide the basis for the perceived distinction between "da" and "ga"; at higher levels, they serve this same phonetic purpose, but also evoke nonspeech whistles. As we found from our own listening, the phonetic information is provided over a range of approximately 20 db below the duplexity threshold;[2] the whistles, which are, of course, barely audible at the duplexity threshold, become louder as the intensity of the sinusoid is further increased. These results show that processing of the sinusoid as speech has priority, thereby defining what we mean by precedence of the phonetic module.

Unlike the earlier form of duplex perception, which required that the transitions and the remainder of the pattern be presented to different ears, the one reported here puts all parts of the pattern equally into both. It thereby avoids such complications of interpretation as may arise with dichotic stimulation, and so makes more straightforward the inference we would draw: that duplex perception reflects distinct auditory and phonetic ways of perceiving the same stimulus.

---

[1] Below the duplexity threshold, such matching would presumably be at chance. Still, it is possible that forced matching is a more sensitive measure than the one we used to obtain the threshold itself. So we applied the matching procedure at 4 db below the lower ("d") threshold, using eight highly practiced subjects. As expected, the responses (45.3% correct, $t(7) = -1.28, p > 0.2$) were at chance.

[2] Bentin & Mann (1983) found a similar range in a dichotic task, though they interpreted it as a difference in sensitivity, not as preemption.

Beyond that, the results obtained with the new form of the duplex phenomenon support the hypothesis that the phonetic mode has prior claim on the transitions, using them for its special linguistic purposes until, having appropriated its share, it passes on the remainder to be perceived by the nonspeech system as "auditory" whistles. Such precedence reflects the profound biological significance of speech.

## References

Bentin, S., & Mann, V. A. (1983). Selective effects of masking on speech and nonspeech in the duplex perception paradigm. *Haskins Laboratories Status Report on Speech Research, SR-76,* 65-85.

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review, 81,* 348-374.

Darwin, C. S., & Sutherland, N. S. (1984). Grouping frequency components of vowels: When is a harmonic not a harmonic? *Quarterly Journal of Experimental Psychology, 36A*, 193-208.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14*, 211-235.

Mattingly, I. G., & Liberman, A. M. (in press). Specialized perceiving systems for speech and other biologically significant sounds. In G. .M. Edelman, W. E. Gall, & W. E. Cowan, (Eds.), *Functions of the auditory system.* New York: Wiley.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology, 2*, 131-157.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception *Psychological Review, 85*, 172-191.

Stevens, K. N. (1975). The potential role of property detectors in the perception of consonants. In G. Fant & M. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 303-330). New York: Academic Press.

# EVIDENCE OF TALKER-INDEPENDENT INFORMATION FOR VOWELS*

**Robert R. Verbrugge**† and **Brad Rakerd**‡

**Abstract.** *The vowel information present in initial and final regions of /b/-vowel-/b/ syllables was examined in this study. Vowels were identified for unedited syllables spoken by a man and a woman, for the initial 20% of those syllables, for the final 20% of the syllables, for the initial and final 20% of the syllables combined and separated by a 60% silent gap, and for the initial and final 20% of the syllables interchanged across talkers and separated by a 60% silent gap. Results indicate: (1) that there is considerable vowel information present in the dynamic regions at the beginnings and endings of syllables; (2) that the information is, to a large extent, carried relationally by those regions; (3) that the information is talker-independent in form; and (4) that the information is complementary to, and distinct from, formant frequency information present in a syllable's center. An experiment assessing the perceived source(s) of these stimuli suggests that source perception is influenced by as yet unspecified acoustic modulations defined at the syllable level.*

## INTRODUCTION

When a vowel is coarticulated with preceding and following consonants to form a syllable, the resulting acoustic pattern usually includes periods of rapid spectral change at its beginning and end, and a period of relative spectral constancy at its center. It is well established that the configuration of formant frequency values present, or best approximated, at the syllable center provides information about the identity of the vowel (e.g., Joos, 1948; Ladefoged, 1975; Peterson & Barney, 1952). After Strange, Jenkins, and Johnson (1983), we will refer to the ideal form of this configuration as an **acoustic target**.

There have been recurring indications that vowel information is also provided by the more dynamic regions of the syllable (Lehiste & Meltzer, 1973; Lindblom & Studdert-Kennedy, 1967;

---

* *Language and Speech*, 1986, *29*, 39-57.

† Now at AT&T Bell Laboratories

‡ Now at Michigan State University

*Acknowledgment.* This research was supported by grants HD01994 and RR05596 to Haskins Laboratories. We are grateful to Winifred Strange, James Jenkins, and David Williams for their many helpful comments on this project.

Shankweiler, Verbrugge, & Studdert-Kennedy, 1978; Strange, Verbrugge, Shankweiler, & Edman, 1976). Perhaps the most compelling evidence of this comes from the experiments of Strange et al. (1983; also Jenkins, Strange, & Edman, 1983). Those investigators assessed the perception of stimuli that preserved only the dynamic beginnings and endings of /b/–vowel–/b/ syllables, the syllable centers having been deleted and replaced with silence. Listeners spontaneously integrated the initial and final portions of these "silent-center" syllables, typically hearing a single utterance with an interruption in the middle (somewhat like a glottal stop). More importantly, vowel identification for these syllables was remarkably accurate, not differing significantly from the accuracy of identification for unedited syllables.

Two competing explanations for this silent-center finding provide the motivation for the present study. First, it is conceivable that listeners used the dynamic regions of those syllables to extrapolate to the formant-frequency targets that had been excised from the syllable centers. Lindblom (1963; also Lindblom & Studdert-Kennedy, 1967) has suggested that listeners make such extrapolations as a matter of course when processing natural speech. Whenever a talker speaks rapidly or destresses the production of a syllable, formant frequencies are "reduced," i.e., they fail to reach target values at the syllable center (Joos, 1948; Lindblom, 1963). Lindblom's (1963) proposal is that in these situations listeners draw on information in the dynamic regions to compute the missing targets. Specifically, they are said to draw on the fact that the initial and final formant trajectories form exponential functions that decelerate toward, or accelerate from, asymptotic target frequencies. To summarize, on this view the dynamic regions of a syllable contribute to vowel perception by subserving the more accurate estimation of target values approximated at the syllable center.

An alternative view of the silent-center result is that the dynamic regions convey vowel information that is complementary to, and distinct from, target information. One way to motivate this alternative is to think of vowels as articulatory events, that is, as gestures that manifest a characteristic organization of forces over the articulators (Fowler, 1977, 1980; Fowler, Rubin, Remez, & Turvey, 1980). From this perspective, the vowels of a dialect are distinguished by different "styles" of articulatory movement. The resulting acoustic modulations provide substantial information about vowel identity, information that differs in kind from the target information present at a syllable's center.

To test the competing claims of the target-extraction and event-perception hypotheses, we constructed *hybrid* silent-center syllables, pairing the initial and final portions of corresponding syllables spoken by a man and a woman. According to the target hypothesis, a hybrid syllable should be very disruptive perceptually. Because the man and woman have different vocal tract sizes and shapes, their corresponding syllable portions should "point to" very different targets. This is illustrated in Figure 1. On the left are spectrograms of the man's and woman's productions of the syllable /bæb/. On the right those spectrograms have been cross-spliced to juxtapose their centers. It is clear that the center formant frequencies are quite discrepant, making it highly unlikely that any extrapolated target values could coincide across talkers.

According to the event hypothesis, a discrepancy in syllable centers is not necessarily disturbing. Talkers who speak a common dialect would be expected to produce a vowel with a common style of articulatory and acoustic change that is independent of idiosyncratic differences in vocal tract size. Therefore, the event hypothesis, in its strongest form, predicts that the woman's and

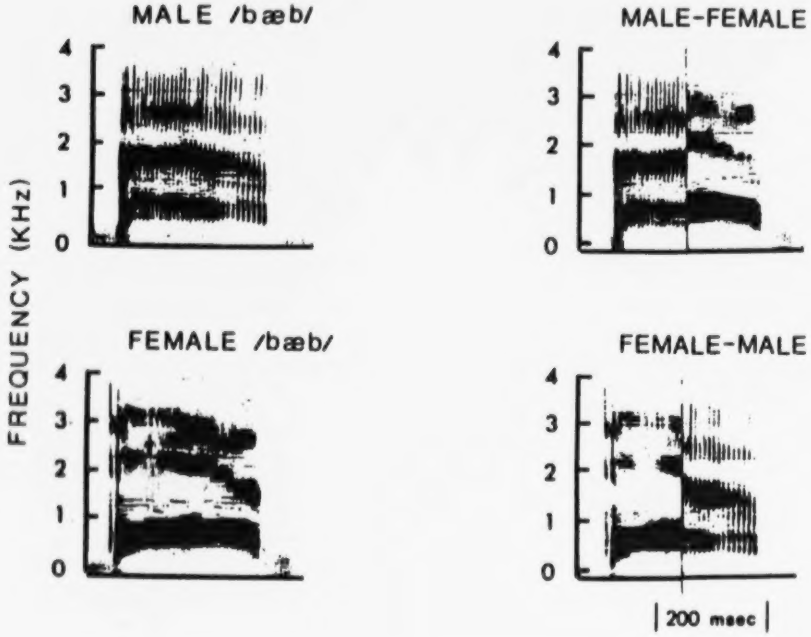

**Figure 1.** Spectrograms of the man's and woman's productions of /bæb/ are presented on the left of the figure. To create the patterns on the right, those spectrograms were cut at the center of their voiced regions, and the initial and final halves were interchanged.

man's syllable portions should be integrated perceptually, and that accuracy of vowel identification should be high, perhaps as high as for single-talker silent-center syllables.

## EXPERIMENT 1: VOWEL PERCEPTION

In this experiment we assessed the accuracy of vowel identification for hybrid-silent-center syllables and for a number of comparison syllables.

### Method

### Stimuli

The stimuli for all experimental conditions were derived from natural speech tokens of /b/-vowel-/b/ syllables. Syllable vowels were the American English vowels /i, ɪ, e, ɛ, æ, ɑ,ʌ,ɔ, o, ʊ, u/). A man and a woman each produced three tokens of each syllable. The syllables were produced in citation form and were paced to match the beat of a metronome. Productions were recorded on audio tape and then digitized for editing (sampling rate = 20 kHz). For each of the eleven vowels, we selected the pair of syllables, one from each talker, that were most closely matched in duration. In general it proved possible to find a very close match. The largest durational disparity was 20 ms and the average disparity was 4.5 ms (2% of the duration of the average voiced region, which was the same for both talkers).

**Table 1**

**The Woman's (W) and Man's (M) Formant Frequency Value in Hz, and Their Absolute Differences Expressed as a Ratio of the Man's Values.**

| | First Formant | | | Second Formant | | | Third Formant | | |
|---|---|---|---|---|---|---|---|---|---|
| Vowel | W | M | (W-M)/M | W | M | (W-M)/M | W | M | (W-M)/M |
| i | 320 | 320 | 0.00 | 2480 | 2080 | 0.19 | 3240 | 2840 | 0.14 |
| ɪ | 400 | 480 | 0.17 | 2080 | 1760 | 0.18 | 2840 | 2480 | 0.15 |
| e | 320 | 400 | 0.20 | 2240 | 1920 | 0.17 | 3000 | 2560 | 0.17 |
| ɛ | 560 | 480 | 0.17 | 1840 | 1520 | 0.21 | 2560 | 2480 | 0.03 |
| æ | 640 | 560 | 0.14 | 2080 | 1480 | 0.41 | 2920 | 2480 | 0.18 |
| ɑ | 720 | 560 | 0.29 | 1320 | 1160 | 0.14 | 2920 | 2480 | 0.18 |
| ʌ | 640 | 480 | 0.33 | 1240 | 1080 | 0.15 | 3000 | 2480 | 0.21 |
| ɔ | 640 | 480 | 0.33 | 1240 | 1000 | 0.24 | 2920 | 2480 | 0.18 |
| o | 480 | 400 | 0.20 | 1000 | 920 | 0.09 | 2760 | 2320 | 0.19 |
| ʊ | 480 | 400 | 0.20 | 1160 | 1000 | 0.16 | 2760 | 2400 | 0.15 |
| u | 320 | 320 | 0.00 | 1160 | 840 | 0.38 | 2760 | 2160 | 0.28 |
| MEAN | | | 0.18 | | | 0.21 | | | 0.17 |
| /e, o/ excluded | | | 0.18 | | | 0.23 | | | 0.17 |

**Table 2**

**Average Women's (W) and Men's (M) Formant Frequency Values in Hz, and Their Absolute Differences Expressed as a Ratio of the Men's Values. These Data Are from Peterson and Barney (1952).**

| | First Formant | | | Second Formant | | | Third Formant | | |
|---|---|---|---|---|---|---|---|---|---|
| Vowel | W | M | (W-M)/M | W | M | (W-M)/M | W | M | (W-M)/M |
| i | 310 | 270 | 0.15 | 2790 | 2290 | 0.22 | 3310 | 3010 | 0.10 |
| ɪ | 430 | 390 | 0.10 | 2480 | 1990 | 0.25 | 3070 | 2550 | 0.20 |
| ɛ | 610 | 530 | 0.15 | 2330 | 1840 | 0.27 | 2990 | 2480 | 0.21 |
| æ | 860 | 660 | 0.30 | 2050 | 1720 | 0.19 | 2850 | 2410 | 0.18 |
| ɑ | 850 | 730 | 0.16 | 1220 | 1090 | 0.12 | 2810 | 2440 | 0.15 |
| ʌ | 760 | 640 | 0.17 | 1400 | 1190 | 0.18 | 2780 | 2390 | 0.16 |
| ɔ | 590 | 570 | 0.04 | 920 | 840 | 0.10 | 2710 | 2410 | 0.12 |
| ʊ | 470 | 440 | 0.07 | 1160 | 1020 | 0.14 | 2680 | 2240 | 0.20 |
| u | 370 | 300 | 0.23 | 950 | 870 | 0.09 | 2670 | 2240 | 0.19 |
| MEAN | | | 0.15 | | | 0.17 | | | 0.17 |

**Spectral comparison 1: Between talkers.** The formants of the woman's vowels (*W* vowels) were typically higher in frequency than the formants of the corresponding vowels spoken by the man (*M* vowels). Table 1 reports their formant frequency values and shows that, on average, those values differed by 18%, 23%, and 17% for the first ($F_1$), second ($F_2$), and third ($F_3$) formants respectively.[1] For comparison, we determined the average formant frequency differences between men and women based on Peterson and Barney's (1952) normative vowel data. That analysis is summarized in Table 2. Peterson and Barney found that formant values of an average adult female talker ($n = 28$) differed from those of an average male talker ($n = 33$) by 15% for $F_1$, 17% for $F_2$, and 17% for $F_3$. The formant frequency differences between the two talkers of the present study were very close to these norms.

**Spectral comparison 2: Within talkers.** In absolute terms, the average formant frequency differences between our *W* and *M* vowels were 80 Hz for $F_1$, 282 Hz for $F_2$, and 420 Hz for $F_3$. We wondered how these values compared with within-talker differences for the production of different vowels. Table 3 shows an analysis in which each talker's formant frequencies were rank-ordered and the differences between neighboring frequencies computed. The average differences were 24, 124, and 68 Hz respectively for $F_1$, $F_2$, and $F_3$ of *M* vowels, and 40, 148, and 68 Hz for $F_1$, $F_2$, and $F_3$ of *W* vowels. All of these values were less than half the size of between-talker production differences. We expect, therefore, that if a listener extrapolated to target values from the beginnings and endings of hybrid syllables, those targets would often be associated with *different* vowels.

The same expectation is supported by an analysis of the distribution of the two talkers' vowel tokens in $F_1$-$F_2$ space. Figure 2 shows that distribution, for a space in which the axes have been scaled to agree with those chosen by Peterson and Barney (1952). Note that for eight of the eleven vowel categories the man's token is closest to a token of a different vowel in the woman's space. In her case the mismatch is even more extreme; 10 of her 11 tokens lie nearest to a token of a different category in his space. This clearly indicates that the initial and final portions of a hybrid syllable would generally "point to" different target vowels when referred against a single talker's $F_1$-$F_2$ space.

### Experimental Conditions

The *W* and *M* syllables were edited for presentation in our experimental conditions according to the general procedures outlined by Strange et al. (1983). Each syllable was divided into three portions. (1) The **initial** portion of a syllable included the release burst of its initial /b/ plus 20% of the voiced region. (2) The **central** portion included the middle 60% of the voiced region. (3) The **final** portion included the final 20% of voicing plus the closure and release of the syllable-final /b/.[2] All measurements were made to the nearest zero-crossing of the speech

---

[1] These figures are based on measurements of the nine vowels for which Peterson and Barney (1952) provide a comparison (/i, ɪ, ɛ, æ, ɑ, ʌ ,ɔ,ʊ. u/). When we include in our analysis the vowels /e, o/, the woman's vowel formant frequencies differ from the man's by an average of 18%, 21%, and 17% for $F_1$, $F_2$, and $F_3$, respectively.

[2] Our editing procedures differed from those of Strange et al. (1983) and Jenkins et al. (1983) in terms of the percentage of the voiced region assigned to initial, center, and final syllable portions. Our choice of 60% as the center proportion is larger, on average, than their value, which varied from 50-60% depending on vowel category. As a result, our silent-center and hybrid-silent-center conditions involve a more severe deletion of signal.

**Table 3**

**The Woman's and Man's Formant Frequencies ($F$) in Hz**
**Rank Ordered and Differenced ($F$-$F$prev).**

| | First Formant | | Second Formant | | Third Formant | |
|---|---|---|---|---|---|---|
| Talker | *F* | *F-Fprev* | *F* | *F-Fprev* | *F* | *F-Fprev* |
| *woman* | 320 | | 1000 | | 2560 | |
| | 320 | 0 | 1160 | 160 | 2760 | 200 |
| | 320 | 0 | 1160 | 0 | 2760 | 0 |
| | 400 | 80 | 1240 | 80 | 2760 | 0 |
| | 480 | 80 | 1240 | 0 | 2840 | 80 |
| | 480 | 0 | 1240 | 0 | 2920 | 80 |
| | 560 | 80 | 1840 | 600 | 2920 | 0 |
| | 640 | 80 | 2080 | 240 | 2920 | 0 |
| | 640 | 0 | 2080 | 0 | 2920 | 0 |
| | 640 | 0 | 2240 | 160 | 3000 | 80 |
| | 720 | 80 | 2480 | 240 | 3240 | 240 |
| MEAN | | 40 | | 148 | | 68 |
| *man* | 320 | | 840 | | 2160 | |
| | 320 | 0 | 920 | 80 | 2320 | 160 |
| | 400 | 80 | 1000 | 80 | 2400 | 80 |
| | 400 | 0 | 1000 | 0 | 2480 | 80 |
| | 400 | 0 | 1080 | 80 | 2480 | 0 |
| | 480 | 80 | 1160 | 80 | 2480 | 0 |
| | 480 | 0 | 1480 | 320 | 2480 | 0 |
| | 480 | 0 | 1520 | 40 | 2480 | 0 |
| | 480 | 0 | 1760 | 240 | 2480 | 0 |
| | 560 | 80 | 1920 | 160 | 2560 | 80 |
| | 560 | 0 | 2080 | 160 | 2840 | 280 |
| MEAN | | 24 | | 124 | | 68 |

waveform. Various combinations of the syllable portions were used to prepare the stimuli for five experimental conditions, as illustrated in Figure 3.

**Whole syllables.** For the whole-syllable condition, all three syllable portions were presented in their original temporal relation (i.e., the syllables were unedited). An example of a whole syllable, the woman's /bæb/, is shown at the top of Figure 3. There were 22 whole-syllable stimuli, 11 different syllables produced by each of the two talkers. These syllables are comparable to the "Control" syllables of Strange et al. (1983) and Jenkins et al. (1983).

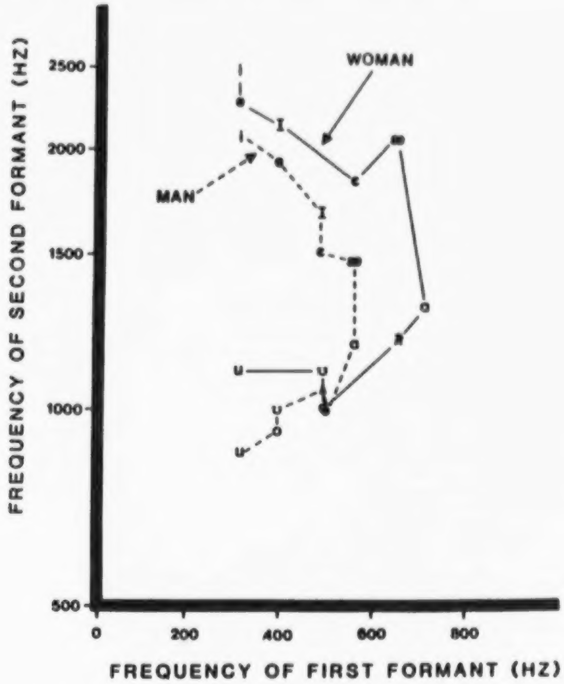

**Figure 2.** Distribution of the man's and woman's vowels in an $F_1/F_2$ space.

**Silent centers.** Second from the top is an example of a silent-center syllable. The central portion of the woman's /bæb/ has been excised and replaced by silence in this instance. We created one silent-center version of each of the 22 syllables.

**Hybrid silent centers.** Third from the top of the figure is an example of a hybrid-silent-center syllable combining the initial portion of the woman's ($W$) /bæb/ with the final portion of the man's ($M$) /bæb/. The silent interval separating these portions was the same as for the woman's silent-center /bæb/. Eleven $W/M$ and 11 $M/W$ hybrids comprised the stimuli of this condition.

*Initial portions.* The 22 initial syllable portions provided the materials for this condition.

*Final portions.* The 22 final syllable portions provided the materials for this condition.

### Subjects

The subjects of this study were undergraduates enrolled in an introductory psychology course. Their participation partially fulfilled a course requirement. All of the subjects were native speakers of English. They had no known hearing difficulties and they had no knowledge of the hypotheses under test. The subjects were randomly assigned to one of the five experimental conditions, distributed as follows: whole-syllable condition ($n$ = 10), silent centers ($n$ = 15), hybrid silent centers ($n$ = 12), initial portions ($n$ = 11), final portions ($n$ = 11).

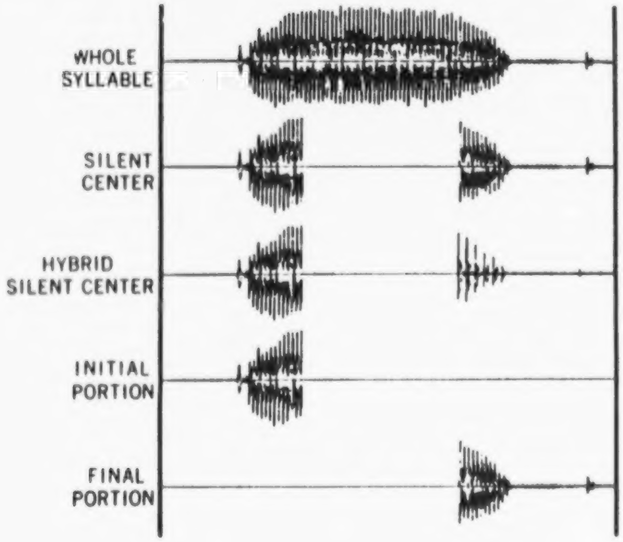

**Figure 3.** Sample tokens of stimuli from the five experimental conditions as indicated. All stimuli derive from the woman's and man's productions of /bæb/.

## Procedure

Stimuli were presented through headphones at a comfortable listening level. A separate group of listeners judged the stimuli of each condition. The subjects were told that they would be hearing edited versions of natural speech, and that they were to decide which of 11 alternative vowels best matched the vowel that they heard on each trial. Their decisions were reported by circling one of 11 /b/–vowel–/b/ words written in English orthography.

Prior to testing, the subjects listened to a demonstration sequence and then completed a block of practice trials. The demonstration sequence consisted of two randomized presentations of the 22 whole-syllable stimuli with two-second pauses between them. The practice block consisted of two randomized presentations of the 22 stimuli of the condition to be tested, with four-second pauses between them. The subjects were required to make responses to the practice stimuli so that they would become familiar with the answer sheet; they were given no feedback as to the accuracy of those responses.

After the practice block, the subjects were allowed to ask questions of clarification about the testing procedure. The testing session commenced immediately after these questions. There were a total of 220 test trials, 10 randomized presentations of the 22 stimuli for a condition. A four-second pause separated succeeding stimuli. Subjects were given a five-minute break halfway through the test.

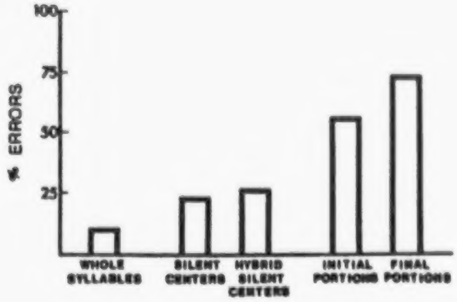

**Figure 4.** Vowel identification error rates for the five experimental conditions. Errors are pooled over 11 vowels and over two talkers.

## Results and Discussion

The overall results for the five listening conditions are displayed in Figure 4. Each bar denotes the mean percentage of errors in vowel identification for the indicated condition, where an error was defined as a failure to categorize a vowel in the same way that the talker intended. Mean percentage errors by condition were as follows: whole syllables (8.8%), silent centers (23.1%), hybrid silent centers (27.4%), initial portions (56.4%), final portions (73.8%). Analysis of variance showed the differences in error rates across conditions to be highly significant: $F(4,54)=144.6$; $p < 0.001$. Post hoc tests (Newman-Keuls) revealed that all pairwise differences among the conditions were significant ($p < 0.01$) with one exception: There was no statistically significant difference between the silent-center and hybrid-silent-center conditions ($p > 0.05$).

### Comparison with Previous Silent-center Studies

Our results replicate and extend the central finding of previous studies examining silent-center stimuli (Jenkins et al., 1983; Strange et al., 1983)—namely, that subjects can identify vowels with good accuracy when syllable centers are silenced. Our results also replicate the previous finding that vowel perception is poor when either initial syllable portions or final portions are presented alone. These results imply, on the one hand, that the dynamic beginnings and endings of syllables are a rich source of information about the syllable vowel and, on the other, that the information is somehow conveyed *relationally* by those beginnings and endings.

One contrast with past studies is our observation of a significant difference between the silent-center and whole-syllable conditions. Previous investigators found no differences between these two conditions (Jenkins et al., 1983; Strange et al., 1983). We may have found a difference in this study because, on average, we deleted a somewhat greater portion of the signal in our silent-center condition than was deleted by others (see footnote 2). Other possible explanations are that there were between-study differences in familiarization with the materials, or in other aspects of the training, or in the subject populations themselves. The overall error rates for both our whole-syllable and silent-center conditions were higher than those seen in previous studies, indicating that others were operating much nearer to the error "floor."

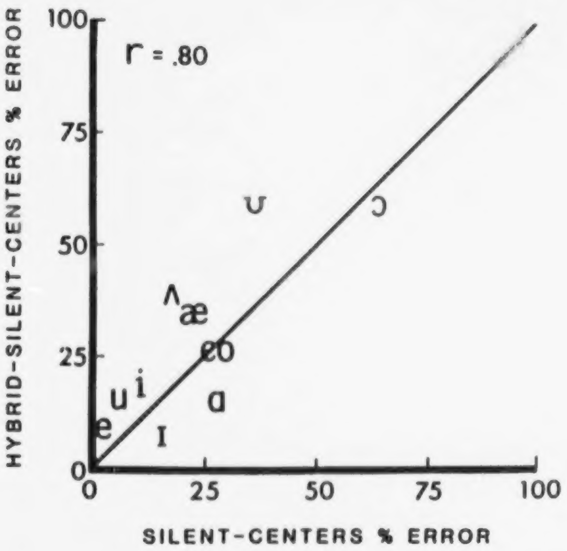

**Figure 5.** Scatter plot of errors for 11 vowels presented in silent-center (abscissa) and hybrid-silent-center (ordinate) conditions. Silent-center errors are collapsed across two talkers, hybrid-silent-center errors are collapsed across man-woman and woman-man hybrids. The coefficient of correlation ($r$) is also provided.

## Silent Centers vs. Hybrid Silent Centers

Of greatest interest to us was the finding that the hybrid and silent-center conditions did not differ significantly. This strongly suggests that the vowel information preserved in silent center syllables is also preserved in hybrids, despite their change of source. That suggestion is strengthened further by a vowel-by-vowel comparison of errors made in the silent-center and hybrid conditions. The comparison is illustrated in Figure 5. Plotted on the abscissa are errors for the silent-center syllables (collapsed over talkers) and on the ordinate are errors for the hybrid syllables (collapsed *M/W* and *W/M* versions). The two data sets are highly correlated ($r = 0.80$; $p < 0.01$), and the clustering of points about the diagonal of the figure demonstrates how similar the errors are in absolute terms. The implication of all of these results is that the vowel information in dynamic regions of a syllable is largely invariant across talkers. It is highly unlikely that this dynamic information subserves the perceptual extraction of any sort of acoustic target, since targets are highly variant across talkers. It is much more likely that the information is indicative of a characteristic articulatory style that is common to productions of the same vowel by talkers of the same dialect.

## The Role of Syllable Duration

Following others (Jenkins et al., 1983; Strange et al., 1983), we have proposed that the dynamic information for vowels is carried relationally by the initial and final syllable portions. Perhaps the simplest relation that might carry it is a durational one. One could imagine that information about the duration of the syllable as a whole could help a listener to distinguish between spectrally-similar, durationally-different vowels in the syllable nucleus. Two lines of evidence speak against this hypothesis. The first comes from a previous study (Strange et al., 1983) that included conditions in which durational differences among silent-center syllables were neutralized. In one condition all of the silent intervals were set equal to the shortest silent duration in the test set and in another they were set equal to the longest. Neither manipulation significantly affected the outcome when all stimuli were produced by a single talker. The "lengthening" manipulation did produce a small but significant increase in errors when different talkers' syllables were interspersed; however, this increase was manifest for vowels of all categories, not just for the short vowels, suggesting that factors other than vowel duration were affected. Overall, there was very little evidence that durational differences among silent-center syllables are an important source of vowel information.

Very little evidence of this can be found in our own results as well. If duration were a primary factor, one would expect the lower error rates for silent centers and hybrids, relative to the initial and final syllable portions, to be due primarily to a reduction in short-long vowel confusions. Short-long vowel errors would be high for the isolated portions (where only spectral information is available), and low for the silent centers, because these syllables presumably supply the duration information needed to distinguish between spectrally-similar short and long vowels.

The first row of Table 4 provides a summary of errors for four spectrally-similar, durationally-different pairs of monophthongs, for each condition of Experiment 1. The second row of the table presents overall errors for the eight vowels after short-long confusions have been removed. The third row summarizes the errors specifically due to short-long confusions. With respect to the duration hypothesis, two observations seem important. First, by the strong form of this hypothesis, errors on isolated portions are due primarily to short-long ("duration") confusions, and overall errors should therefore be roughly *equal* for silent centers and for the isolated portions after duration errors have been removed. The data in Table 4 (second row) do not support this prediction. Second, while more duration errors are observed for isolated portions than for silent centers (third row), the *proportion* of errors attributable to short-long confusions stays relatively constant across these conditions (see fourth row of the table). This suggests that the silent-center format *does not* differentially reduce duration-based errors, but has a broader, and different, kind of impact in reducing perceptual errors. Parametric studies using a broader set of stimulus materials will be needed to address this question further.

## Modeling the Relationship between Initial and Final Portions

If the initial and final syllable portions are not affording listeners a better estimate of intrinsic vowel duration, then how is it that perception is so much better in the silent-center and hybrid-silent-center conditions? One might argue that it is better because in these conditions listeners are, in effect, given two chances to identify the vowel, one chance based on the initial portion and a second based on the final portion. In this section we consider this alternative.

How might initial-portion and final-portion percepts be processed to derive a single vowel judgment? The simplest possibility is that those percepts are, for each vowel, perfectly independent and that a listener simply chooses between them at random. If so, we would expect that errors in the silent-center and hybrid conditions should average 71% (the mean of initial- and final-portion error rates). A nonrandom selection process could, at best, produce error rates of 56% (taking the better of the initial- and final-portion rates for each vowel). Even the latter prediction is much higher than the actual error rates observed for silent centers (23%) and hybrids (27%). Moreover, it poorly predicts the ordering of error rates across vowels: The correlation

**Table 4**

**Mean Percentage Errors on Eight Vowels, /ɪ, i, ɛ, æ, ʌ, ɑ, ʊ, u/, Including and Excluding Confusions on Adjacent Short-long Vowels**

| | Whole Syllable | Silent Center | Hybrid Silent Center | Initial Portion | Final Portion |
|---|---|---|---|---|---|
| Overall errors | 8.3 | 20.1 | 26.3 | 47.9 | 65.9 |
| Overall errors, excluding short-long errors[a] | 3.5 | 11.8 | 17.8 | 26.3 | 39.7 |
| Short-long errors | 4.8 | 8.3 | 8.5 | 21.6 | 26.2 |
| Proportion[b] | 0.58 | 0.41 | 0.32 | 0.45 | 0.40 |

[a] Short-long errors are confusions within any one of the following four vowel pairs: /ɪ-i/, /ɛ-æ/, /ʌ-ɑ/, /ʊ-u/.

[b] Short-long vowel errors as a proportion of overall errors.

between the nonrandom guessing prediction and the observed errors for silent centers was 0.41, and for hybrids it was 0.42.

One might propose a more sophisticated decision model in which the initial- and final-portion percepts are processed in contingent fashion to arrive at a vowel response. For example, the initial portion could be used to narrow down the set of alternatives and the final portion to make a selection from among this reduced set. A good candidate for the initial classification is the intersection of two major phonetic dimensions: high-vs.-low and front-vs.-back. With respect to this four-way classification, listeners made an average of 26% errors when categorizing vowels in the initial-portion condition (excluding the diphthongs /e, o/). Estimates of the probability for error when making the final selection within these categories can be derived from our data on the final portions. When the probabilities for error in the two stages are combined, one obtains a predicted error rate for judgments on the silent-center and hybrid syllables as a whole.[3] Figure 6 shows the comparison between predicted and observed errors for the hybrid condition (the silent-center comparison looks similar). Like the previous models, this contingent model generally overpredicts the absolute level of errors and poorly predicts the patterning of errors

---

[3] For example, in the final-portion condition the high-back vowels /ʊ/ and /u/ were confused with one another on 8% (/ʊ-u/) and 34% (/u-ʊ/) of trials. These percentages, in combination with the probability of making an error when categorizing a high-back vowel as high-back in the initial-portion condition (20%), provided our contingent estimates for /ʊ/ and /u/.

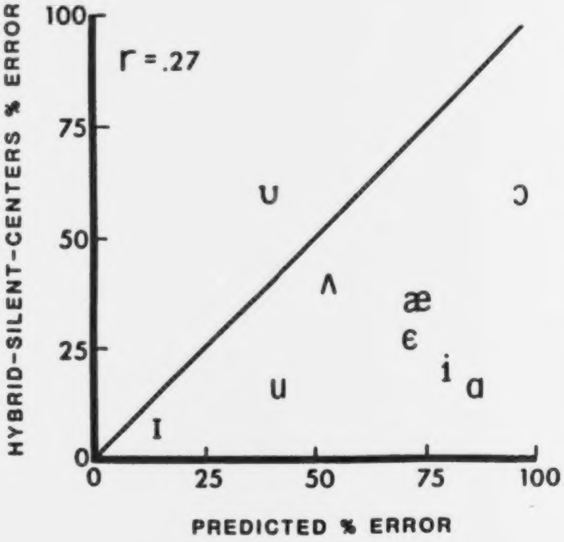

**Figure 6.** Scatter-plot (and correlation) of errors on hybrid silent-center syllables, as predicted by a contingent-judgment model (abscissa) and as observed in the identification test (ordinate).

among the vowels. The correlations between predicted and observed errors were 0.27 for hybrid vowels and 0.50 for silent-center vowels.

The models we tested all assumed that the syllable portions were analyzed separately, and were only related at a late stage in a decision process. This type of perceptual analysis would seem to be demanded by the target-extraction view, which proposes that on-glide and off-glide functions separately specify a target. In particular, separate perceptual analyses would seem to be the only way the target view could approach the perception of *hybrid* syllables, since the syllable portions specify very different asymptotic targets in this case. However, all of the "separate analysis" models underpredict listeners' accuracy on the hybrids by a wide margin. This strongly suggests that a listener does not process the syllable portions separately but, instead, derives vowel information from a "superadditive" relation between them. In other words, it suggests that some singular function over the two portions of a hybrid is detected by the perceiver as the basis for a vowel judgment.

This account of the hybrid-syllable results is compatible with the event hypothesis, which holds that the early and late stages of an event should bear a principled relation to one another. Defining such relations in acoustic terms is a major challenge for future research. The simplest possibility is to define a duration measure over the hybrid syllable as a whole. However, neither our results nor those of Strange et al. (1983) provide much support for syllable duration as the critical "superadditive" relation (see previous section). More complicated possibilities involve characteristic frequency and amplitude modulations over the syllable. Whatever function we may discover, our hybrid data suggest that it will be talker-independent in form, and that it will not be the sum of two exponentials sharing a common asymptote.

The judgment models also raise questions about the role of more local sources of information for vowel identity. The contingent model, for example, considered the possibility that different regions of a syllable provide different kinds of information. While that particular model proved uninformative, there was some evidence in listeners' errors on the isolated portions that the early and late regions of a syllable carry some information about vowel properties. Listeners in both conditions showed better-than-chance performance (chance would be 91% errors). Also, as we noted above, the initial portions of the monophthongs carried sufficient information to support four-way classification (high-low, front-back) with only 26% errors. A similar analysis of errors on final portions shows 33% errors for the four-way classification.

These results on the isolated portions raise a second challenge for future research: to identify the carriers of information in these more local regions of a syllable. In the case of the initial portions, one candidate is the release burst of initial stop consonants. In fact, several studies have reported that this brief initial phase of a syllable is sufficient for better-than-chance discrimination within small sets of vowels (Blumstein & Stevens, 1980; Winitz, Scheib, & Reeds, 1972). The acoustic basis for these effects is still not clear, nor is it clear how well listeners could do on a larger, more representative set of vowels. Even so, these findings provide a good example of a general principle we seek to develop in this paper: The transient regions of a syllable may provide information that is specific to a vowel without necessarily being information about a target state. A rough analogy can be drawn to the role of onset transients in the identification of musical instruments. The dynamic structure of these transients carries more information about instrument identity than does the steady-state region of a sustained tone (Grey & Gordon, 1978; Luce & Clark, 1967; Saldanha & Corso, 1964). More to the point, the transients do not simply aid the extraction of steady-state timbre; they provide information that is different in kind. In the case of vowels, we expect to find a similar pattern: namely, that the structure of a talker's onset transients is both specific to the vowel and distinct from spectral targets.

## EXPERIMENT 2: SOURCE PERCEPTION

After the completion of each session of testing in Experiment 1, we informally interviewed subjects about their impressions of the edited syllables and were surprised to discover that subjects in the hybrid condition rarely heard a complete change of source. Instead, they heard a single talker, typically a male, and, more particularly, a male prone to abrupt pitch changes. These reports were surprising because the hybrid stimuli contain marked discontinuities of fundamental and formant frequencies, and these would normally be expected to specify a change of articulatory source. The perceptual reports suggest that the hybrid syllables contain other types of acoustic information, which strongly specify a single production by a single source. In the normal course of events, this acoustic structure would parallel other information about the source, such as fundamental frequency and formant frequency contours. However, in the unique case of the hybrid stimuli, it opposes these other sources of information and appears to predominate over them. Since this speculation has implications for the study of source perception, we thought it important to make a more rigorous test of the findings that prompted it. In Experiment 2, we directly sought subjects' judgments of the number of talkers they heard when listening to hybrid silent-center (and silent-center) stimuli.

## Method

### Subjects

Nine undergraduate students were the subjects of this experiment. They were native speakers of English with normal hearing. They had no contact with the subjects of Experiment 1 and were not themselves subjects of that experiment.

### Stimuli

The stimuli of this experiment were the silent-center and hybrid-silent-center stimuli described in Experiment 1.

### Procedure

Ten randomized repetitions of the 22 silent-center stimuli, spaced at four-second intervals, comprised a silent-center test block. A comparable arrangement of the 22 hybrid stimuli comprised a hybrid test block. Each test block was presented to subjects twice, in alternation. Five subjects began with the silent-center block, four began with the hybrid block. The subjects' task was to determine which of the following three alternatives best described the source(s) of the stimuli heard on each trial: (1) One talker speaking with normal intonation; (2) One talker speaking with a pitch change; (3) Two talkers speaking. Responses were reported by checking off the appropriate alternative on an answer sheet.

Prior to testing, subjects completed a practice block in which they responded to one presentation of each hybrid and silent-center stimulus. The order of these presentations was randomized. The subjects received no feedback regarding the accuracy of their responses. The practice block was followed by a pause for questions regarding procedure, and then by the first test block. There was a five-minute break between the test blocks. All testing was completed in a single session.

## Results and Discussion

The results of this experiment are summarized in Figure 7, which shows the proportion of silent-center and hybrid responses in each category (collapsed across the two orders of presentation). The results confirm the informal reports given by subjects in the hybrid condition of Experiment 1: Hybrid stimuli are most generally perceived to have been produced by a single talker. They were so perceived on a total of 75% of the trials in the present experiment. That percentage was only slightly smaller than the total percentage of single-talker responses for silent-center syllables (82%). The principal difference between hybrid and silent-center responses was in their distribution over the two single-talker categories. With silent-center stimuli, subjects more often judged that the talker spoke with normal intonation (57% of all judgments), while with hybrid stimuli, subjects more often heard a pitch change (43% of all judgments).

Listeners' judgments that the hybrid stimuli derived from a single source may have been facilitated by the presence of the silent gap between the initial and final portions spoken by the different talkers. The stimuli did not contain *instantaneous* changes in fundamental frequency and formant contours. Instead, those contours were heard to be interrupted at one point and resumed at another. Perhaps in such cases it is reasonable for listeners to ascribe the gap's "bridge" to the rather curious behavior of a single talker. If so, we would note that there is a stong asymmetry

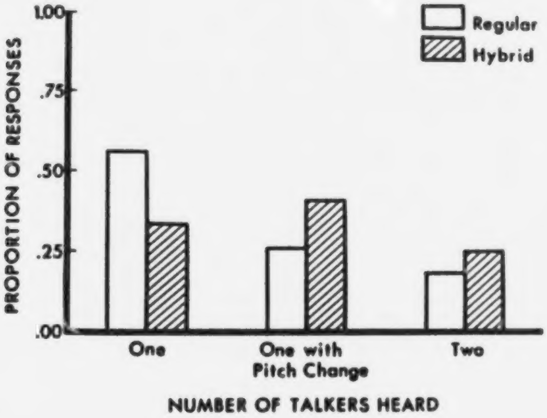

**Figure 7.** Proportion of trials on which subjects judged regular (single-talker)and hybrid silent-center stimuli to have been produced by: (1) one talker speaking with a normal pitch; (2) one talker speaking with an abrupt pitch change; or (3) two talkers.

in those ascriptions. With both woman/man and man/woman hybrids, listeners nearly always reported that the single talker was a man. For some reason his vocal characteristics predominated.

We would also note that few perceptual gaps can be "bridged" so readily as the hybrid gap. While a pitch break of the magnitude seen across the initial and final portions is conceivable for a single talker, a formant pattern break of the magnitude seen (15-20%) is inconceivable. (It would require a change in the talker's age or sex in mid-utterance.) Listeners integrated the syllable portions in spite of this radical change in effective vocal tract dimensions, and this suggests that other, more powerful information for source continuity was present in the acoustic signal. It seems likely that listeners were strongly aided in bridging the silent gap by the common style with which the two talkers produced the original syllables. The two talkers spoke the same dialect and produced the same vowel gestures, in the same phonetic context, under the same timing regimen (matching the beats of a metronome). The close similarity of their articulatory styles would produce, as a natural consequence, a close similarity of acoustic "styles of change" in their productions. These dynamic consequences of "producing the same vowel with the same timing" may be the basis for subjects' integrating the two portions perceptually and hearing them as the product of a common source. Given the composition of the hybrid syllables, we can conclude that this acoustic information is defined over the syllable as a whole, and, in particular, that it is defined sufficiently by a relation between the initial and final regions of the syllable.

## CONCLUSION

Experiments 1 and 2 provide strong indications that the perception of vowel identity and source continuity is sensitive to dynamic acoustic structure defined over the course of a whole syllable. The acoustic information appears to be distinct in type from such variables as syllable duration and spectral targets (whether realized in the signal or extrapolated). Vowel perception and source perception can be remarkably impervious to discontinuities in local spectrum, if speech materials are otherwise matched in timing and articulatory style. This strongly suggests that a dialect's vowels can be characterized by higher-order variables (patterns of articulatory and spectral *change*) that are independent of a specific talker's vocal tract dimensions. A more precise definition of these variables will aid our understanding of the acoustic basis for identifying a vowel and, not coincidentally, for perceiving an articulation as continuous.

## References

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America, 67*, 648-662.

Fowler, C. A. (1977). *Timing control in speech production*. Doctoral dissertation, University of Connecticut. (Reproduced by the Indiana University Linguistics Club, Bloomington, Indiana.)

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8*, 113-133.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production* (pp. 373-420). New York: Academic Press.

Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America, 63*, 1493-1500.

Jenkins, J. J., Strange, W., & Edman, T. R. (1983). Identification of vowels in "vowelless" syllables. *Perception & Psychophysics, 34*, 441-450.

Joos, M. (1948). Acoustic phonetics. *Language Supplement, 24*, 1-136.

Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Javonovich.

Lehiste, I., & Meltzer, D. (1973). Vowel identification in natural and synthetic speech. *Language and Speech, 16*, 356-364.

Lindblom, B. E. F. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America, 35*, 1773-1781.

Lindblom, B. E. F., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *Journal of Acoustical Society of America, 42*, 830-843.

Luce, D., & Clark, M. Jr. (1967). Physical correlates of brass-instrument tones. *Journal of the Acoustical Society of America, 42*, 1232-1243.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175-184.

Saldanha, E. L., & Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America, 36*, 2021-2026.

Shankweiler, D. P., Verbrugge, R. R., & Studdert-Kennedy, M. (1978). Insufficiency of the target for vowel perception. *Haskins Laboratories Status Report on Speech Research, SR-55/56*, 103-111.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America, 60*, 213-224.

Strange, W., Jenkins, J. J., & Johnson, T. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America, 74*, 694-705.

Winitz, H., Scheib, M. E., & Reeds, J. A. (1972). Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech. *Journal of the Acoustical Society of America, 51*, 1309-1317.

# CONTROLLED VARIABLES IN SENTENCE INTONATION*

**Carole E. Gelfer,† Katherine S. Harris,† and Thomas Baer**

## INTRODUCTION

In describing the acoustic characteristics of sentence intonation, the terms *downdrift* and *declination* have been applied to the behavior of both the rapid variations in fundamental frequency ($F_0$) corresponding to syllable prominences whose peaks comprise the envelope of an $F_0$ contour (see, for example, Cooper & Sorenson, 1981), and the slower variation in $F_0$ that defines a reference level upon which these local prominences are superimposed (see, for example, Cohen, Collier, & 't Hart, 1982). Recently, there has been considerable interest in the mental representation of various aspects of declination (Breckenridge, 1977; Cooper & Sorenson, 1981; Liberman & Pierrehumbert, 1982; Pierrehumbert, 1979) and, by extension, the control or regulation of the physiological variables involved in its realization (Atkinson, 1973; Collier, 1975; Gelfer, Harris, Collier, & Baer, 1985; Maeda, 1976). Unfortunately, cognitive processes are not readily observable. However, to the extent that they are expected to have some *physical* reality, examining the patterns of control of the physiological processes that ultimately bear on the acoustic aspects of sentence intonation should provide some insight into the *psychological* reality of declination.

In the first part of this paper, we will examine the behavior of subglottal pressure ($P_s$) during speech in order to determine whether the time course of the drop in subglottal pressure associated with declination is a controlled variable in sentence intonation, or, alternatively, the passive consequence of lung deflation. Obviously, the rate at which air is used in producing speech depends on the phonetic characteristics of utterances (Klatt, Stevens, & Mead, 1968). For example, because of the reduced airflow resistance at the glottis and the configuration of the vocal tract for a voiceless fricative, substantially higher airflow rates occur for utterances containing the syllable /fa/ than for those containing syllables composed of voiced continuants, such as /ma/. If the lungs were allowed to deflate passively, we would expect subglottal pressure to decline at different rates over the course of these syllables. However, there is evidence indicating that lung deflation during speech is not a purely passive phenomenon. For example, Draper, Ladefoged, and Whitteridge (1960) and Mead, Bouhuys, and Proctor (1968) found subglottal pressure to be stable throughout sustained voice production, thus suggesting that the muscles of the respiratory system are marshalled in such a way as to maintain $P_s$. However, these studies have examined only sustained phonations of constant amplitudes that also require constant pressures. On the

---

* In T. Baer, C. Sasaki, & K. Harris (Eds.), *Laryngeal function in phonation and respiration* (pp. 422-435). Boston: College-Hill Press, 1987.

† Also City University of New York

*Acknowledgment.* This work has been supported by NINCDS Grants 13870 and 13617 and BRS Grant RR-05596 to Haskins Laboratories.

other hand, subglottal pressures during speech are known to vary dynamically. What we do not know, then, is whether the variation in pressure over time is the natural by-product of unchecked expiratory forces, or whether it reflects ongoing control of the respiratory musculature in order to produce dynamically stable pressures. By using reiterant speech (Kelso, V-Bateson, Saltzman, & Kay, 1985; Larkey, 1983) in which a sentence is mimicked with a high flow syllable, "fa," or a low flow syllable, "ma," we can discover whether the time course of pressure variation is the by-product of unchecked expiratory forces, or whether it is dynamically stable. Moreover, to the extent that $F_0$ mirrors $P_s$, we can perhaps gain insight into the factors responsible for declination itself.

In the second part of this paper, we will address the phenomenon known as $F_0$ resetting. It has been suggested that the declination function is sensitive to the syntactic structure of an utterance. Thus, in a two-clause utterance, the $F_0$ contour may be discontinuous at the major syntactic boundary so that a single falling contour no longer characterizes the declination function (Cooper & Sorenson, 1981; Fujisaki & Hirose, 1982; Maeda, 1976). However, there is some question as to which aspect of the $F_0$ trajectory actually defines resetting in these instances. For example, Fujisaki and his colleagues (Fujisaki & Hirose, 1982; Fujisaki, Hirose, & Ohta, 1979) have developed a model of intonation that allows for two basic inputs, the phrase level and accent level commands, which are realized as the 'voicing' (baseline) and 'accent' (syllabic) components, respectively. According to this model, it is the voicing component that may be reset at clause boundaries, while the accent components vary independently of the baseline, and, therefore, independently of syntactic structure.

Cooper and Sorenson (1981) suggest, too, that declination is reset at clause boundaries in a way that is relevant to the syntactic structure of an utterance. However, in contrast to Fujisaki, they measure declination, and thus gauge resetting, on the basis of the relationship of syllable peaks; specifically, the height of the first peak in a second clause to that of a sentence-initial peak. Furthermore, they suggest that the resetting of peak $F_0$ directly mirrors a speaker's intention to signal the syntactic structure of the sentence, and that resetting is planned in some detail at the outset of an utterance. While we recognize that there is an interaction between syntax and the realization of sentence intonation, we hypothesize that the *extent* to which $F_0$ is reset is *not* planned prior to the execution of an utterance even if the presence or absence of resetting may be planned. Fujisaki has suggested that resetting is triggered when a significant pause occurs at the clause boundary. Taking this notion a step further, we would suggest instead that it is not only the pause but also the new inspiration that may accompany it that in turn influences $F_0$ indirectly through the resetting of such variables as subglottal pressure and/or laryngeal muscle activity. Thus, we hypothesize that $F_0$ resetting will depend on the presence or absence of a pause and inspiration at clause boundaries.

## METHODS

Two speakers served as subjects for the first part of this study, and one of the two served as a subject for the second. Both are native speakers of Dutch, fluent in English, and both were aware of at least some of the purposes of this work. They were chosen as subjects primarily because of their willingness, and ability, to tolerate the invasive procedures required.

Lung volume was inferred from the calibrated sum of thoracic and abdominal signals from a Respitrace inductive plethysmograph, and airflow rate (cc/sec) was derived from calculations of volume over time. Subglottal pressure was recorded directly, but differently, for the two subjects, RC and LB. For RC, a pressure transducer (Setra Systems 236L) was coupled to the subglottal space by means of a cannula inserted percutaneously through the cricothyroid membrane. For LB, a miniature pressure transducer (Millar SPC-350) was introduced pernasally through the posterior glottis into the trachea. While the percutaneous approach is certainly the more invasive procedure, it provides a signal that is easier to calibrate, because the miniature transducer cannot be calibrated outside the body, and it is highly sensitive to changes in temperature that occur within the trachea upon inspiration (Cranen & Boves, 1985). Unfortunately, we did not recognize these difficulties at the time of recording, so that the pressure signal could not be calibrated properly. However, while absolute values for the pressure data for the subject using this device are uninterpretable, the relative pressure levels should be valid, since temperature changes affect the zero offset but not the sensitivity of the transducer. For both subjects, EMG techniques previously described (Harris, 1981) were used to record from the cricothyroid muscle. Fundamental frequency was derived from the output of an accelerometer (Stevens, Kalikow, & Willemain, 1975) attached to the pretracheal skin surface. For LB, a cepstral technique was used to extract $F_0$ from the signal. For RC, the accelerometer output was sampled using a Visipitch period-by-period $F_0$ extractor. This latter procedure is equal in accuracy to the former $F_0$ extraction technique, but has the advantage of on-line sampling at one-half real time. However, it became available to us only after the data for the first subject had been analyzed.

### Stimuli

In the first experiment, the two subjects produced reiterant forms of Dutch utterances, using the syllables /ma/ and /fa/ (Appendix A). These utterances were also produced in three lengths, with three different emphatic stress configurations (early, double, and late). Thus, there were nine utterance types per reiterant condition (i.e., /ma/ or /fa/). However, the stress and length conditions will not be discussed separately here, except to be noted in the examples shown, because the differences among them have been discussed previously (Gelfer et al., 1985).

In the second experiment, one of the subjects, RC, produced three similar English sentences. For two of the sentences, the syntactic boundary was moved in order to alter slightly the length of each clause. The third sentence conjoined two clauses similar to those comprising the first two sentences (Appendix B). The subject's task was to produce each sentence under two conditions: no pause and no inspiration at the clause boundary, and both a pause and an inspiration at the clause boundary.

## RESULTS: EXPERIMENT 1

Averaged subglottal pressure, lung volume, and the amplitude envelope for utterances of Length 2 with various emphatic stress configurations are shown for both subjects in Figure 1. It is apparent from this figure that, for the subglottal pressure, there is little difference between the /ma/ and /fa/ utterances apart from the presence of local perturbations in the curve of the /fa/ utterances. The acoustic amplitude envelopes of the two reiterant utterance types show no substantial difference in overall acoustic amplitude, and, as would be expected, resemble the subglottal pressure contours in overall shape. However, despite the uniformity of the pressure curves, the lung volume curves for the two utterances show the change in volume over time to be

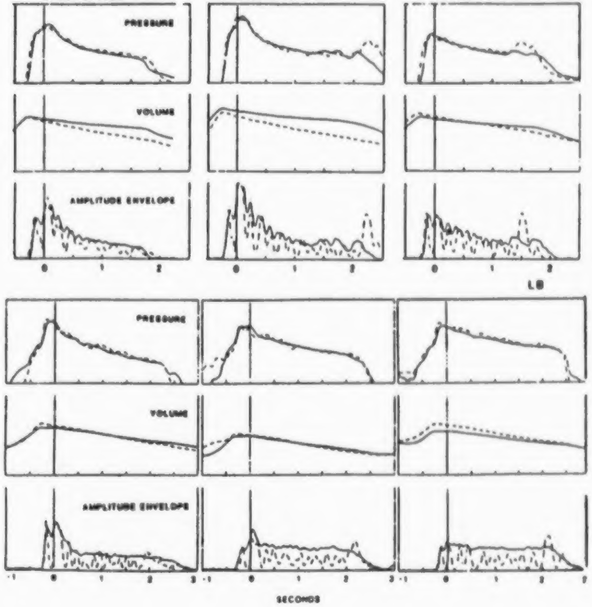

**Figure 1.** Averaged subglottal pressure (panel 1), Respitrace (panel 2), and amplitude envelope (panel 3) curves for comparable /ma/ and /fa/ utterances for subjects RC (top) and LB (bottom). The vertical line in each panel denotes the line-up point used for averaging the tokens of each utterance type, which in these utterances is the onset of the vowel for the first syllable receiving lexical stress. The solid curves represent the reiterant /ma/ utterances, and the dashed curves the reiterant /fa/ utterances. The maximum and minimum values for pressure on the y axis are 13 cm $H_2O$ are 0 cm $H_2O$ for RC, 9 cm $H_2O$ and -6 cm $H_2O$ for LB. For respiratory valence, values range from 5 liters to 2 liters for RC, and from 5 liters for 1 liter for LB. The audio amplitude is in arbitrary units.

greater for the /fa/ utterances, as is evidenced by the steeper slopes. Thus, for both subjects, we observe no apparent relationship between airflow rate and the $P_s$ contours.

In order to quantify these data, we plotted the distributions of subglottal pressures and airflow rates for the two utterance types. For subglottal pressure, we measured average levels over a fixed time interval, rather than differences over time, in order to neutralize any segmental effects. Since our earlier work demonstrated that effects of such variables as sentence length are reflected in initial peak pressure values (Gelfer et al., 1985), we were careful to eliminate these portions of the curves from the measured interval. By calculating the averages over an interval of 600 ms, from 400 to 1000 ms, after the occurrence of the first lexically stressed syllable, we were able to avoid averaging values under these peaks, at the same time being able to include data from some of the shortest utterances.

The same interval was used to calculate the change in lung volume over time. However, because the Respitrace curves are rather smooth and not prone to perturbation due to segmental effects, we calculated the difference in volume between the two points in order to derive the rate of decline (i.e., airflow rate).

The distributions of $P_s$ measures for all tokens of the /ma/ and /fa/ utterances are shown in Figure 2. The difference between the means of these distributions is statistically nonsignificant: $p > .2$ for RC; $p > .5$ for LB. By contrast, the difference in airflow rate for the /ma/ and /fa/ utterances (Figure 3) *is* statistically significant for both subjects, $p < .001$. Thus, $P_s$ appears to remain stable despite the significant differences in airflow secondary to the phonetic structure of these utterances.

## RESULTS: EXPERIMENT 2

In this experiment, Subject (RC) produced three two-clause utterances under conditions where pausing and inspiration were directly manipulated. In the first condition, he produced each repetition of each utterance with neither a pause nor inspiration at the clause boundary. In the second condition, all tokens were produced with both a pause and inspiration at the clause boundary.

Figure 4 shows the averaged Respitrace and $P_s$ curves for one sentence across the two conditions being considered here (i.e., −pause/−inspiration and +pause/+inspiration). This general picture is identical across sentence types, so we will present graphic displays only for one sentence.

In the absence of both a pause and inspiration at the clause boundary in the first condition (Panel 1), there is a continuous, although choppy, subglottal pressure curve throughout both clauses and across the intervening boundary as well. On the other hand, where both a pause and inspiration occur (Panel 2), there is a concomitant drop in the subglottal pressure during the inspiration, which then increases significantly as expiration resumes.

Despite the differences in pause durations and respiratory activity, the subject produced the same general $F_0$ contours across conditions (Figure 5). For our analyses, $F_0$ values were measured for the first peak in the first clause (peak 1A), the last peak in the first clause (peak 1B), and the first peak in the second clause (peak 2A) for the five tokens of each of the three sentences under each condition.

Figure 6 is a schematic respresentation of the average values, collapsed across sentence type, for each condition. It can be seen that, while the $F_0$ values for the two peaks (1A and 1B) in the first clause are strikingly similar across conditions, the value of the first peak in the second clause (2A) varies systematically as a function of the pausing/breathing condition at the clause boundary. That is, where there is no pause or inspiration, $F_0$ falls 8 Hz below those peaks that were preceded by an inspiration (Table 1). This difference is statistically significant as well, $p < .001$.

A comparison of $P_s$ values at peak 2A yields corresponding results. That is, subglottal pressure is significantly higher when a pause and inspiration occur than when they do not, $p < .001$. Moreover, when the ratio of frequency change per centimeter of water is calculated for peak 2A between conditions 1 and 2, these ratios fall within the accepted range of 3-7Hz/Cm-$H_2O$ (Baer, 1979; Hixon, Klatt, & Mead, 1971; Ladefoged, 1963), suggesting that the relationship between the increase in $P_s$ and that in $F_0$ could be more than a correlational one. However, before the behavior of $F_0$ is attributed to the presence or absence of an increase in $P_s$, the contribution of laryngeal muscle activity must be determined.

Figure 7 shows the cricothyroid muscle activity for the two conditions for the same sentence. It appears that there is no systematic resetting of CT activity as a function of inspiration at

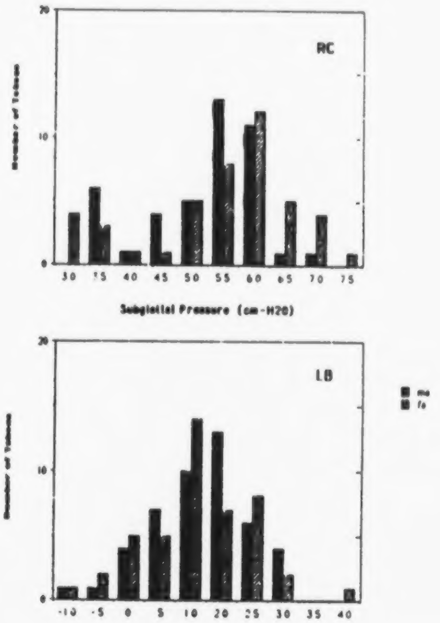

**Figure 2.** Distribution of $P_s$ averages for tokens of all /ma/ and /fa/ utterances for both subjects. The solid bars denote the /ma/ tokens, and the dashed bars the /fa/ tokens.

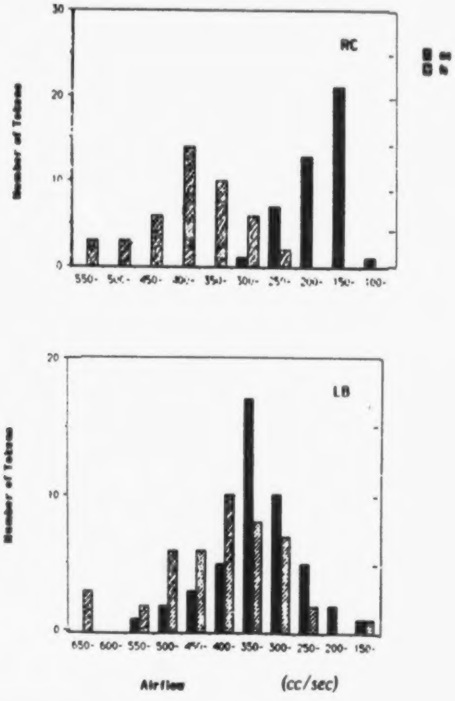

**Figure 3.** Distribution of airflow rates (cc/sec) for tokens of all /ma/ and /fa/ utterances for both subjects. The solid bars denote the /ma/ tokens, and the dashed bars the /fa/ tokens.

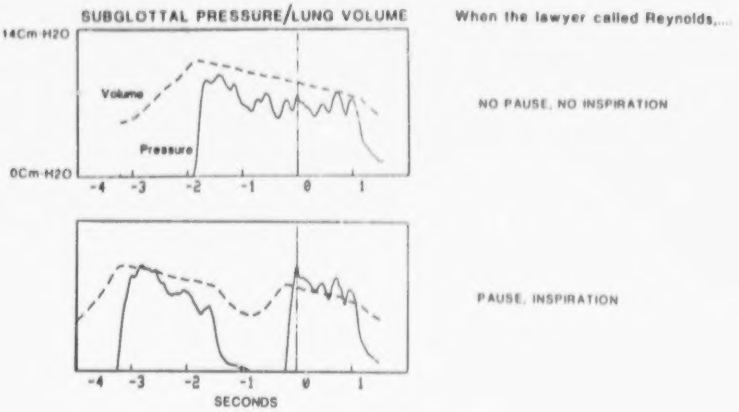

**Figure 4.** Averaged subglottal pressure and Respitrace curves for a representative sentence across conditions. The first panel represents the no pause, no inspiration condition, and the second panel represents the pause plus inspiration condition. The line-up point, depicted by the vertical line, represents the onset of voicing for the vowel in the word 'plan' in the second clause. The same line-up point was used for all three sentence types.

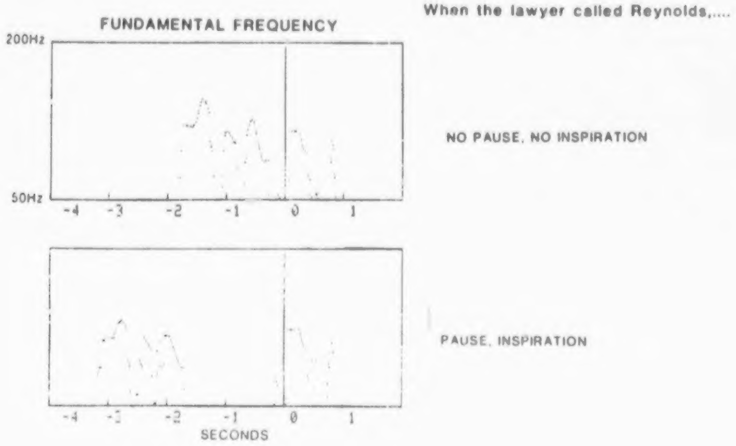

**Figure 5.** Averaged $F_0$ contours for a representative sentence across conditions. The first panel represents the no pause, no inspiration condition, and the second panel represents the pause plus inspiration condition. The line-up point, depicted by the vertical line, represents the onset of voicing for the vowel in the word 'plan' in the second clause. The same line-up point was used for all three sentence types.

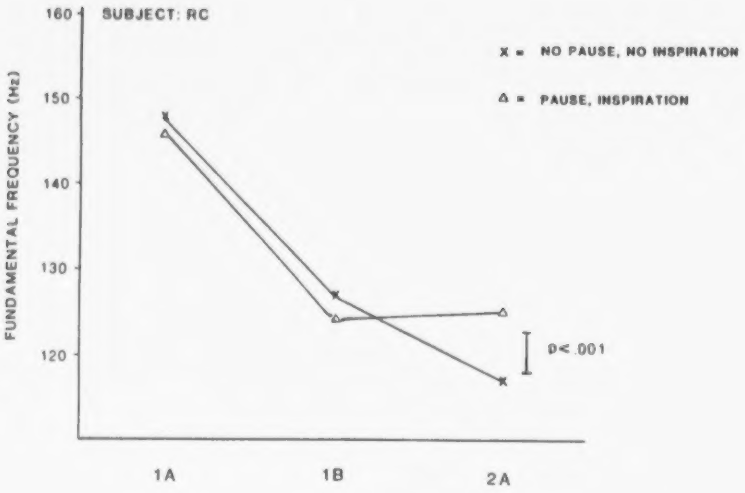

**Figure 6.** Schematic representation of the mean $F_0$ values (peaks 1A, 1B, 2A), collapsed across sentence types, for both conditions. The X's denote the no pause, no inspiration condition, and the triangles's the pause plus inspiration condition.

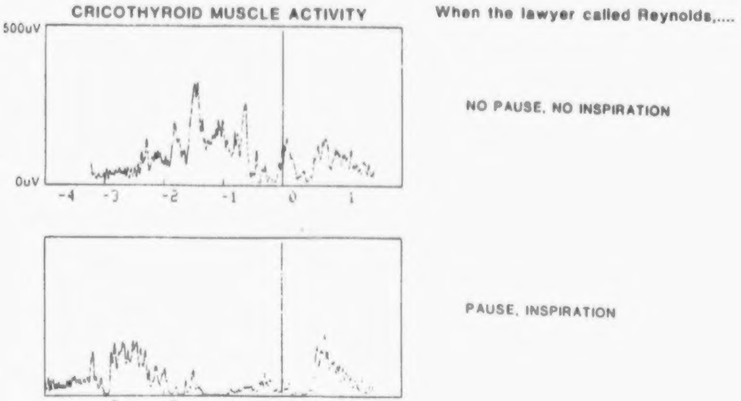

**Figure 7.** Averaged cricothyroid muscle activity for a representative sentence across conditions. The first panel represents the no pause, no inspiration condition, and the second panel represents the pause plus inspiration condition. The line-up point, depicted by the vertical line, represents the onset of voicing for the vowel in the word 'plan' in the second clause. The same line-up point was used for all three sentence types.

**Table 1**

**Values at Peak 2A**

| | **Fundamental Frequency** | | **Subglottal Pressure** | |
|---|---|---|---|---|
| | **−Pause/−Insp** | **+Pause/+Insp** | **−Pause/−Insp** | **+Pause/+Insp** |
| Sentence 1 | 116 | 123 | 7.8 | 9.9 |
| Sentence 2 | 117 | 122 | 8.6 | 10.1 |
| Sentence 3 | 118 | 130 | 9.1 | 11.6 |
| Mean | 117 | 125 | 8.5 | 10.5 |

| | **Condition 2 - Condition 1** | | **Ratio** |
|---|---|---|---|
| | **$F_0$** | **$P_s$** | **Hz/Cm-$H_2O$** |
| Sentence 1 | 7 | 2.1 | 3.33 |
| Sentence 2 | 5 | 1.5 | 3.33 |
| Sentence 3 | 12 | 2.5 | 4.80 |

Averaged $F_0$ and $P_s$ values for peak 2A for the three sentences and the ratios of Hz/Cm-$H_2O$ calculated between the no pause/no inspiration and the pause plus inspiration conditions.

the clause boundary. In fact, there is more CT activity following the clause boundary in the first condition, where no inspiration occurs. It would thus appear that CT contributes little, if any, to $F_0$ resetting in this case, and that the increase in $P_s$ following an inspiration could indeed account for the amount of resetting observed. The above results suggest that when both a pause and inspiration occur, there is a significant increase in $P_s$ and $F_0$ values relative to those occurring when there is neither a pause nor an inspiration. However, in comparing only these two conditions, we are unable to separate the relative effects of breathing and pausing on resetting.

Our results differ somewhat from those of Collier (1987) who, in certain instances, found a greater amount of resetting. In addition, Collier fails to find the substantial effect of inspiration on $P_s$ that we do. We believe that these differences may be attributed to differences in the tasks in the two studies. That is, while Collier manipulates the stress configuration (i.e., lo-lo; hi-hi) around the clause boundary, we do not. Thus, the intentional realization of specific intonation contours might result, for example, in greater involvement in CT activity while, at the same time, reducing $P_s$ activity.

## Discussion

It has been known for some time that the respiratory system acts in such a way as to stabilize subglottal pressure (eg., Draper et al., 1960; Mead et al., 1968). The data presented here not only confirm the results of these earlier studies, but provide evidence that this control is dynamic in nature. Furthermore, this stability is maintained even when the system must respond to perturbations in the form of varying airflow requirements. In other words, if lung deflation were passive in nature, pressure would certainly decline more rapidly for utterances where greater airflow rates are used. However, we have found the rate of pressure decline to be independent of the rate of airflow.

Previous studies in which simultaneous measures of subglottal pressure and fundamental frequency have been recorded during sentence production have noted that, through the most stable portions of these curves, their decline is relatively parallel (see, for example, Atkinson, 1973; Collier, 1975; Lieberman, 1967), although a direct cause and effect relationship has been difficult to establish. However, Gelfer et al. (1985) were able to demonstrate that, in the absence of cricothyroid activity, the fall in pressure accounted for an appropriate fall in frequency. Moreover, the rate of both $P_s$ and $F_0$ decline was found to be stable across varying utterance lengths. The data presented here suggest that $P_s$ is a controlled variable in sentence production, and that $F_0$ declination is a consequence.

Similarly, the resetting of $F_0$ at a clause boundary appears to represent the effect of a general resetting of the respiratory system on subglottal pressure following an inspiration. That is, we found $F_0$ to be significantly higher when an inspiration occurred at the clause boundary than when it did not. At the same time, however, it is difficult to make the claim that the resulting difference of 8 Hz is perceptually salient, for it is also the case that the syntactic structure can be easily recovered when listening to any token of any of these utterances. It is not entirely clear, then, that peak $F_0$ resetting is a necessary mechanism for encoding syntactic structure on the part of the speaker, or a prerequisite for decoding syntax on the part of the listener. Furthermore, that the extent of $F_0$ resetting is planned by a speaker, in that it has a place in the mental representation of an utterance, seems untenable. Rather, resetting would appear to be the outcome of an optional speaker strategy—perhaps, for example, whether a speaker chooses to pause or take a new breath, and thus "reset" the whole system, prior to the execution of a second clause—and that this is the level at which it is controlled.

## References

Atkinson, J. E. (1973). *Aspects of intonation in speech: Implications from an experimental study of fundamental frequency*. Unpublished doctoral dissertation, University of Connecticut.

Baer, T. (1979). Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. *Journal of the Acoustical Society of America, 65*, 1271-1275.

Breckenridge, J. (1977). Declination as a phonological process. Murray Hill, NJ: Bell System Technical Memorandum.

Cohen, A., Collier, R., & 't Hart, J. (1982). Declination: Construct or intrinsic feature of pitch? *Phonetica, 39*, 254-273.

Collier, R. (1975). Physiological correlates of intonation patterns. *Journal of the Acoustical Society of America, 58*, 249-255.

Collier, R. (1987). $F_0$ declination: The control of its setting, resetting, and slope. In T. Baer, C. Sasaki, & K. Harris (Eds.), *Laryngeal function in phonation and respiration* (pp. 403-421). Boston: College-Hill Press, 1987.

Cooper, W. E., & Sorenson, J. M. (1981). *Fundamental frequency in sentence production*. New York: Springer-Verlag.

Cranen, B., & Boves, L. (1985). Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production. *Journal of the Acoustical Society of America, 77*, 1543-1551.

Draper, M. H., Ladefoged, P., & Whitteridge, D. (1960). Expiratory pressures and air flow during speech. *British Medical Journal, vol. i*, 1837-1843.

Fujisaki, H., & Hirose, K. (1982). Modeling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation. *Preprints of papers, Working Group on Intonation* (pp. 57-70). The XIIIth International Congress of Linguistics, Tokyo.

Fujisaki, H., Hirose, K., & Ohta, K. (1979). Acoustic features of the fundamental frequency contours of declarative sentences in Japanese. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, 13*, 163-173.

Gelfer, C. E., Harris, K. S., Collier, R., & Baer, T. (1985). Is declination actively controlled? In I. Titze (Ed.), *Vocal fold physiology; Physiology and biophysics of the voice*. Iowa City, IA: Iowa University Press.

Harris, K. S. (1981). Electromyography as a technique for laryngeal investigation. In C. L. Ludlow & M. O. Hart (Eds.), *ASHA Reports: Proceedings of the Conference on the Assessment of Vocal Pathology, 11*, 70-86.

Hixon, T. J., Klatt, D. H., & Mead, J. (1971). Influence of forced transglottal pressure changes on vocal fundamental frequency. *Journal of the Acoustical Society of America, 49*, 105.

Kelso, J. A. S., Vatikios-Bateson, E., Saltzman, E. L., & Kay, B. (1985) A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America, 77*, 266-280.

Klatt, D. H., Stevens, K. N., & Mead, J. (1968). Studies of articulatory activity and airflow during speech. In A. Bouhuys (Ed.), *Sound production in man. Annals of the New York Academy of Science*, 42-54.

Ladefoged, P. (1963). Some physiological parameters in speech. *Language and Speech, 6*, 109-119.

Larkey, L. S. (1983). Reiterant speech: An acoustic and perceptual evaluation. *Journal of the Acoustical Society of America, 73*, 1337-1345.

Liberman, M., & Pierrehumbert, J. (1982). Intonational invariance under changes in pitch range and length. *Bell Laboratories Internal Technical Memorandum*.

Liberman, M., & Streeter, L. A. (1978). Use of nonsense-syllable mimicry in the study of prosodic phenomena. *Journal of the Acoustical Society of America, 63*, 231-233.

Lieberman, P. (1967). *Intonation, perception and language*. Research Monograph No. 38. Cambridge: MIT Press.

Maeda, S. (1976). *A characterization of American English intonation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Mead, J., Bouhuys, A., & Proctor, D.F. (1968). Mechanisms generating subglottic pressure. In A. Bouhuys (Ed.), *Sound production in man. Annals of the New York Academy of Science*, 177-181.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America, 66*, 363-369.

Stevens, K. N., Kalikow, D. N., & Willemain, T. R. (1975). A miniature accelerometer for detecting glottal waveforms and nasalization. *Journal of Speech and Hearing Research, 18*, 594-599.

## APPENDIX A

*Early Stress*:

Length 1: Je *weet* dat jan nadenkt.

Length 2: Je *weet* dat jan erover nadenkt te betalen.

Length 3: Je *weet* dat jan erover nadenkt ons darvoor met genoegen te betalen.

*Double Stress*:

Length 1: Je *weet* dat jan *na*denkt.

Length 2: Je *weet* dat jan erover nadenkt te be*ta*len.

Length 3: Je *weet* dat jan erover nadenkt ons daarvoor met genoegen te be*ta*len.

*Late Stress*:

Length 1: Je weet dat jan *na*denkt.

Length 2: Je weet dat jan erover nadenkt te be*ta*len.

Length 3: Je weet dat jan erover nadenkt ons daarvoor met genoegen te be*ta*len.

## APPENDIX B

*Sentence 1*: When the lawyer called Reynolds, the plans were discussed.

*Sentence 2*: When the lawyer called, Reynolds' plans were discussed.

*Sentence 3*: The lawyer called Reynolds, and the plans were discussed.

# ARTICULATORY SYNTHESIS: NUMERICAL SOLUTION OF A HYPERBOLIC DIFFERENTIAL EQUATION

**Richard S. McGowan**

**Abstract.** *The computation of acoustic pressure fluctuations in a variable area tube is often done using the Kelly-Lochbaum reflection model. The numerical scheme derived from this model can be put into the context of finite-difference approximations to a differential equation describing acoustic wave propagation (a hyperbolic differential equation). Quantitative criteria for goodness of finite-difference schemes (truncation error, stability, and dispersion) are discussed without considering the effect of boundary conditions. An alternative scheme that has better truncation error to the reflection model approximation is examined, but we do not necessarily recommend its adoption. The quantitative criteria should be applied to the full initial-boundary value problem inherent in articulatory synthesis when a numerical scheme is being chosen.*

## INTRODUCTION

In this note one aspect of articulatory synthesis will be considered—that of solving the differential equation describing acoustic (small amplitude), one-dimensional propagation of a pressure disturbance through a lossless tube with spatially varying cross section. This equation can be written:

$$\frac{\partial^2 p}{\partial t^2} = cY^{-1}\frac{\partial}{\partial x}\left(cY\frac{\partial p}{\partial x}\right) \tag{1}$$

where $Y(x) = A_0/\rho_0 c$ acoustic admittance, $A_0(x)$ = cross-sectional area of the tube when no disturbances are present, $\rho_0$ = density of air with no disturbances, $c$ = adiabatic speed of sound in air, $p$ = acoustic perturbation pressure, $t$ = time, and $x$ = distance along the tube axis (Lighthill, 1978, pp. 124-125). This equation (Webster horn equation) will be known as the differential equation for the remainder of this note. This equation belongs to the class of hyperbolic differential equations the meaning and consequences of which will be discussed in the rest of this note.

In current articulatory synthesis, both time domain and frequency domain, the Kelly Lochbaum reflection model provides a popular method for the computation of sound propagation in a tube (Liljencrants, 1985; Rubin, Baer, & Mermelstein, 1981). This method can be

---

*Acknowledgment.* Preparation of this manuscript was supported by Grant NS13617 to Haskins Laboratories. The author thanks Philip Rubin for helpful comments.

seen to be a finite-difference approximation to the differential equation. In this context there are quantitative measures for goodness of approximation to the solution of the differential equation. Three of these will be discussed here: stability, truncation error, and dispersion relations. Roughly, a stable method is one for which the solution remains bounded in a finite time span as the discrete time interval goes to zero. Truncation error tells us how much better we would do in approximating the differential equation if we were to reduce the discrete time and discrete spatial intervals of the finite-difference equation. In other words, it says how well the solution to the differential equation solves the finite-difference approximation. Dispersion relations, or the relationship between frequency and wavenumber, can be derived for solutions to both the differential equation and the finite-difference approximation. These relationships should express the same relation to a close approximation because the ratio of frequency to wavenumber gives the phase speed (Trefethen, 1982). Dispersion error has been considered previously in the speech literature, where it is sometimes called frequency warping (Maeda, 1982; Portnoff, 1973).

Considerations of truncation error will allow us to propose another finite-difference approximation, which is a slight modification to that provided by the Kelly-Lochbaum reflection model. Then we will consider the stability and dispersion relations of both approximations. Because the boundary conditions inherent in the articulatory synthesis problem are not considered in the analysis here, we cannot recommend one method over the other. The alternative method illustrates the possibility of deriving other efficient finite-difference schemes with, perhaps, better numerical properties.

We will be applying the von Neumann stability condition to the finite-difference methods in this note. This condition does not provide a sufficient condition for the full initial-boundary value problem of articulatory synthesis. Under special circumstances it does provide necessary and sufficient conditions for pure initial value problems with constant coefficient difference equations (Richtmeyer & Morton, 1967, pp. 68-72). However, the von Neumann condition, applied locally, will be a necessary condition for strong stability in the full initial-boundary value problem (Richtmeyer & Morton, 1967, p. 99, 132). The von Neumann condition will be stated below when it is invoked.

## STRUCTURE OF THE DIFFERENTIAL EQUATION

First, we will explore the structure of the differential equation with a few transformations, which will help illustrate the meaning of the phrase: hyperbolic differential equation. The second order differential equation can be written as a system of two first order differential equations:

$$\frac{\partial p}{\partial t} = -cY^{-1}\frac{\partial J}{\partial x} \tag{2}$$

$$\frac{\partial J}{\partial t} = -cY\frac{\partial p}{\partial x} \tag{3}$$

where $J$ is the perturbation volume velocity in the small amplitude limit (Lighthill, 1978, pp. 124-125). In matrix notation:

$$\mathbf{I}\frac{\partial}{\partial t}\mathbf{U} + c\mathbf{A}\frac{\partial}{\partial x}\mathbf{U} = \mathbf{0} \tag{4}$$

where:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 0 & Y^{-1} \\ Y & 0 \end{pmatrix},$$

$$\mathbf{U} = \begin{pmatrix} p \\ J \end{pmatrix}, \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Next, we will perform a couple of similarity transformations on the $(p, J)$ space that will help to simplify the form of equation (4). The $(p, J)$ space is transformed by a stretching transformation, **B**, and then a rotation, **G**. Because the dependent variables, $p$ and $J$, are transformed, the differential equation (4) must also be transformed. In particular, the coefficient matrix **A** will be transformed to a matrix in diagonal form. Let:

$$\mathbf{V} = \mathbf{GBU}, \tag{5}$$

where

$$\mathbf{B} = \begin{pmatrix} \sqrt{Y} & 0 \\ 0 & \sqrt{Y^{-1}} \end{pmatrix},$$

$$\mathbf{G} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Thus:

$$\mathbf{V} \equiv \begin{pmatrix} v^+ \\ v^- \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{Y^{-1}}J + \sqrt{Y}p \\ \sqrt{Y^{-1}}J - \sqrt{Y}p \end{pmatrix}. \tag{6}$$

As a result of the transformations on the $(p, J)$ space, the system (4) is transformed into (see appendix):

$$\mathbf{I}\frac{\partial}{\partial \mathrm{t}}\mathbf{V} + \mathrm{c}\mathbf{H}\frac{\partial}{\partial \mathrm{x}}\mathbf{V} = \mathrm{c}\mathbf{K}\mathbf{V}, \tag{7}$$

where

$$\mathbf{K} = \begin{pmatrix} 0 & -1/2\frac{dlog(x)}{dx} \\ 1/2\frac{dlog(x)}{dx} & 0 \end{pmatrix},$$

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \mathbf{GBA}(\mathbf{GB})^{-1}.$$

By the form of the relation between them, **A** and **H** are similar matrices. We have diagonalized the coefficient matrix that is multiplying the spatial derivative (i.e., transformed **A** to **H**), while leaving the identity matrix as the coefficient matrix of the time derivative. Because **H** has real eigenvalues, the system (7) is hyperbolic. Because **A** and **H** are similar, **A** has the same two real eigenvalues, and the system (4) is hyperbolic, and the original differential equation (1) is a hyperbolic differential equation.

The implications for a system having the property of being hyperbolic are best illustrated by considering the system (7). By a change of the independent variables in (7), we can make further simplifications. Let:

$$\zeta = t + x/c, \quad \xi = t - x/c. \tag{8}$$

In terms of these variables, system (7) becomes:

$$\begin{aligned}\frac{\partial v^+}{\partial \zeta} &= -\frac{c}{4}\frac{d\,log(Y)}{dx}v^- \\ \frac{\partial v^-}{\partial \xi} &= \frac{c}{4}\frac{d\,log(Y)}{dx}v^+.\end{aligned} \tag{9}$$

Also, equation (8) can be expressed as a set of differential equations:

$$\frac{\partial t}{\partial \zeta} = 1/c\frac{\partial x}{\partial \zeta}, \quad \frac{\partial t}{\partial \xi} = -1/c\frac{\partial x}{\partial \xi}. \tag{10}$$

The set of equations (9) and (10) constitutes the canonical system of the original differential system (4) (Forsythe & Wasow, 1960, p. 43). One way to solve the second-order hyberbolic system in two independent variables is by integrating the system (9) simultaneaously along the characteristic lines, $\zeta = constant$ and $\xi = constant$, given by (10). Because each component equation in system (9) involves derivatives of the dependent variable along one characteristic line only, they may be treated as coupled ordinary differential equations for the sake of computation.

In Figure 1, the geometry of the situation in the $(x/c, t)$ plane is illustrated. For articulatory synthesis, the inflow boundary conditions are normally specified for $x/c = 0$, and impedance boundary conditions at $x/c = l/c$. An initial condition should also be specified at $t = 0$. This leads to a well-posed initial-boundary value problem (Higdon, 1986). The discussion of boundary conditions will be postponed until a later note, and only the pure initial value problem will be considered here.

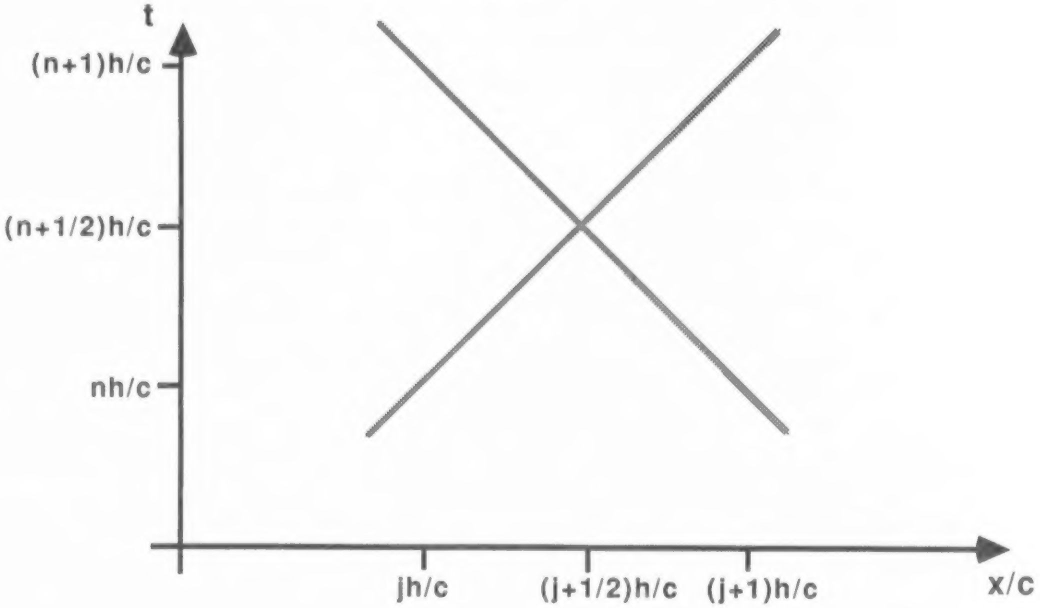

**Figure 1.** Characteristic lines.

The meanings of the superscripts "+" and "−" on the dependent variables $v^+$ and $v^-$ in equation (6) will now be explained. First, express volume velocity, $J$, and pressure, $p$, in terms of volume velocities in the positive and negative x-direction.

$$J = J^+ - J^-, \quad p = Y^{-1}\left(J^+ + J^-\right) \tag{11}$$

It can be seen that the new dependent variables are just scaled versions of the positive-going volume velocity and the negative-going volume velocity. The scaling depends upon spatial position.

$$v^+ = \sqrt{2Y^{-1}}J^+, \quad v^- = \sqrt{2Y^{-1}}J^- \tag{12}$$

Note that in the case of constant area and no boundaries there is a particularly simple solution to (9) and (10) — that of two waves travelling at speed $c$ in opposite directions, without change of form. More generally, if the logarithmic derivative of the area with respect to $x$ is small, then energy along characteristic lines is approximately constant. This is seen by noting that the right-hand sides of (9) are approximately zero, and that the intensities of the positive and negative-going waves are $Y^{-1}(J^+)^2$ and $Y^{-1}(J^-)^2$ respectively (see Lighthill, 1978, pp. 120-123).

## THE REFLECTION MODEL AS A FINITE-DIFFERENCE SCHEME

The Kelly-Lochbaum reflection model can be seen to provide a finite-difference approximation to the system (8) if the following approximations are made. We will make reference to Figure 1, and let the step sizes be defined:

$$\Delta\zeta = \Delta\xi = 2h/c, \qquad \Delta x/c = \Delta t = h/c, \tag{13}$$

where $h > 0$.

We will be assuming that the dependent variables and the admittance function all have continuous third derivatives. This smoothness condition allows us to make use of Taylor's formula with remainder to estimate truncation error in the approximations. Normally, truncation error is written in terms of powers of the step size. For example, $f(x,h)$ is said to approximate $g(x)$ to $\mathcal{O}(h^N)$ if:

$$g(x) = f(x,h) + q(x,h),$$

where:

$$0 < \lim_{h\to 0}\left|\frac{q(x,h)}{h^N}\right| < \infty.$$

We normally write:

$$g(x) = f(x,h) + \mathcal{O}(h^N)$$

However, we will sometimes write the function $q(x,h)$ explicitly to show how the error depends on the smoothness of certain other functions.

The first derivatives in equation (9) are approximated:

$$\begin{aligned}
\frac{\partial v^{+(n+1/2)}_{(j+1/2)}}{\partial \zeta} &= \frac{v^{+(n+1)}_{(j+1)} - v^{+(n)}_{(j)}}{(2h/c)} - \frac{1}{6}\frac{\partial^3 v^+}{\partial \zeta^3}\bigg|_{\zeta=\zeta^*}\left(\frac{h}{c}\right)^2 \\
\frac{\partial v^{-(n+1/2)}_{(j+1/2)}}{\partial \xi} &= \frac{v^{-(n+1)}_{(j)} - v^{-(n)}_{(j+1)}}{(2h/c)} - \frac{1}{6}\frac{\partial^3 v^-}{\partial \xi^3}\bigg|_{\xi=\xi^*}\left(\frac{h}{c}\right)^2,
\end{aligned} \tag{14}$$

where $v^{+(n)}_{(j)}$ refers to $v^+(t,x)$ at $t = nh/c$ and $x/c = jh/c$, and $(j+n)h/c < \zeta^* < (j+n+2)h/c$ and $(n-j-1)h/c < \xi^* < (n-j+1)h/c$. The derivatives are approximated by centered differences along the characteristic lines. The logarithmic derivative of the admittance must also be approximated.

$$\begin{aligned}
\frac{d\,log Y_{(j+1/2)}}{dx} &= \frac{1}{Y_{(j+1/2)}}\frac{dY_{(j+1/2)}}{dx}, \\
&= \frac{2/h\left(Y_{(j+1)} - Y_{(j)}\right) - 1/3\frac{d^3Y}{dx^3}\Big|_{x=x^*}(h/2)^2}{\left(Y_{(j+1)} + Y_{(j)}\right) - \frac{d^2Y}{dx^2}\Big|_{x=x^{**}}(h/2)^2} \\
&= \frac{2}{h}\left(\frac{Y_{(j+1)} - Y_{(j)}}{Y_{(j+1)} + Y_{(j)}}\right) + \mathcal{O}(h^2) \\
&\equiv \frac{2}{h}\mu_{(j+1/2)} + \mathcal{O}(h^2)
\end{aligned} \tag{15}$$

where $jh < x^*, x^{**} < (j+1)h$. The finite-difference approximation to this derivative is, to within a constant factor, the reflection coefficient, $\mu_{(j+1/2)}$, for a tube with a discontinuous change in area at $x = (j+1/2)h$. Note, given the smoothness of $Y(x)$, that:

$$\mu_{(j+1/2)} = \frac{h}{\left(Y_{(j+1)} + Y_{(j)}\right)}\frac{dY}{dx}\bigg|_{x=x^{***}} \leq \mathcal{O}(h),$$

where $jh < x^{***} < (j+1)h$. The values of the dependent variables at $t = (n+1/2)h/c$, $x/c = (j+1/2)h/c$ also need to be estimated:

$$\begin{aligned}
v^{+(n+1/2)}_{(j+1/2)} &= v^{+(n)}_{(j)} + \frac{\partial v^+}{\partial \zeta}\bigg|_{\zeta=\zeta^{**}}\left(\frac{h}{c}\right) \\
v^{-(n+1/2)}_{(j+1/2)} &= v^{-(n)}_{(j+1)} + \frac{\partial v^-}{\partial \xi}\bigg|_{\xi=\xi^{**}}\left(\frac{h}{c}\right),
\end{aligned} \tag{16}$$

where $(n-j-1)h/c < \xi^{**} < (n-j+1)h/c$ and $(n+j)h/c < \zeta^{**} < (n+j+2)h/c$.

We write the resulting finite-difference approximation to (9), using relations (11) through (16):

$$\begin{aligned} J_{(j+1)}^{+(n+1)} &= \sqrt{\frac{Y_{(j+1)}}{Y_{(j)}}}\,J_{(j)}^{+(n)} + \left[-\mu_{(j+1/2)} + \mathcal{O}(h^3)\right]\left[-J_{(j+1)}^{-(n)} + \mathcal{O}(h)\right] + \mathcal{O}(h^3) \\ J_{(j)}^{-(n+1)} &= \sqrt{\frac{Y_{(j)}}{Y_{(j+1)}}}\,J_{(j+1)}^{-(n)} - \left[\mu_{(j+1/2)} + \mathcal{O}(h^3)\right]\left[J_{(j)}^{+(n)} + \mathcal{O}(h)\right] + \mathcal{O}(h^3) \end{aligned} \tag{17}$$

The square roots can be approximated:

$$\begin{aligned} \sqrt{\frac{Y_{(j+1)}}{Y_{(j)}}} &= 1 + \mu_{(j+1/2)} + \mathcal{O}(h^2) \\ \sqrt{\frac{Y_{(j)}}{Y_{(j+1)}}} &= 1 - \mu_{(j+1/2)} + \mathcal{O}(h^2) \end{aligned} \tag{18}$$

Finally, the finite-difference approximation can be written:

$$\begin{aligned} J_{(j+1)}^{+(n+1)} &= (1 + \mu_{(j+1/2)})J_{(j)}^{+(n)} + \mu_{(j+1/2)}J_{(j+1)}^{-(n)} + \mathcal{O}(h^2) \\ J_{(j)}^{-(n+1)} &= (1 - \mu_{(j+1/2)})J_{(j+1)}^{-(n)} - \mu_{(j+1/2)}J_{(j)}^{+(n)} + \mathcal{O}(h^2) \end{aligned} \tag{19}$$

Neglecting the truncation errors, these relations are the same as those provided by the Kelly-Lochbaum model (Markel & Gray, 1976, pp. 66-67). In the analysis presented here, it is necessary that $\frac{dY}{dx}$, $\frac{d^2Y}{dx^2}$, and $\frac{d^3Y}{dx^3}$ be small in order for the truncation error to be small, that is, $Y$ should be relatively smooth.

Another analysis may be possible for a discontinuous admittance function. Work on matched asymptotic expansions has shown that the conditions of continuity of pressure and volume velocity used in the derivation of the Kelly-Lochbaum reflection model is valid to the first order in a compactness parameter, even at abrupt area changes (Lesser & Lewis, 1972). A compactness parameter would be the ratio of the width of the tube section to the wavelength, where the tube width is assumed to be much smaller than the wavelength of sound. This may not be justified if the tube sections are so short that the cut-off modes can leak from one section to another (Thompson, 1984).

Note that an $\mathcal{O}(h^2)$ error is made in the approximation (18). This could be avoided simply by using $v^+$ and $v^-$ as the dependent variables, rather than $J^+$ and $J^-$. $\mathcal{O}(h)$ errors are made in equations (16) in the evaluation of $v_{(j+1/2)}^{-(n+1/2)}$ and in the evaluation of $v_{(j+1/2)}^{+(n+1/2)}$. This error could be improved to be $\mathcal{O}(h^2)$ by taking averages. That is, approximate $v_{(j+1/2)}^{+(n+1/2)}$ by $\frac{1}{2}(v_{(j+1)}^{+(n+1)} + v_{(j)}^{+(n)})$ and $v_{(j+1/2)}^{-(n+1/2)}$ by $\frac{1}{2}(v_{(j)}^{-(n+1)} + v_{(j+1)}^{-(n)})$. If these changes are made, the resulting finite-difference approximation appears as:

$$\begin{pmatrix} v_{(j+1)}^{+(n+1)} \\ v_{(j)}^{-(n+1)} \end{pmatrix} = \frac{1}{1 + (\mu_{(j+1/2)}/2)^2} \begin{pmatrix} 1 - (\mu_{(j+1/2)}/2)^2 & -\mu_{(j+1/2)} \\ \mu_{(j+1/2)} & 1 - (\mu_{(j+1/2)}/2)^2 \end{pmatrix} \begin{pmatrix} v_{(j)}^{+(n)} \\ v_{(j+1)}^{-(n)} \end{pmatrix}. \tag{20}$$

In matrix notation, the approximation provided by the Kelly-Lochbaum model, equation (19), is:

$$\begin{pmatrix} J_{(j+1)}^{+(n+1)} \\ J_{(j)}^{-(n+1)} \end{pmatrix} = \begin{pmatrix} 1 + \mu_{(j+1/2)} & \mu_{(j+1/2)} \\ -\mu_{(j+1/2)} & 1 - \mu_{(j+1/2)} \end{pmatrix} \begin{pmatrix} J_{(j)}^{+(n)} \\ J_{(j+1)}^{-(n)} \end{pmatrix}. \tag{21}$$

### Stability and Dispersion Errors

In the following, we would like to find whether the Euclidian norms of the solution vectors to equations are uniformally bounded in a finite time interval, locally in space, as the step size, $h$, approaches zero (Richtmeyer & Morton, 1967, pp. 68-73). This is a local stability property. Local stability is used since the matrices are functions of the spatial coordinates and without a complete specification of boundary conditions, we cannot talk about the difficult global stability problem. However, to have global stability in the strong sense as defined by Richtmeyer and Morton (1967, p. 99), it is necessary to have the stability defined above in the local sense.

Operationally, the local stability can be determined in the following way. Take a Fourier transform of the dependent variables against the spatial coordinate. Let $y = \exp(\imath kh)$. For example:

$$v_{(j)}^{+(n)} = \int_{-\infty}^{+\infty} \tilde{v}^{+(n)}(k) \exp(\imath(jh)k)dk = \int_{-\infty}^{+\infty} \tilde{v}^{+(n)}(k) y^j dk.$$

For each Fourier component, equations (20) and (21) become:

$$\begin{pmatrix} \tilde{v}^{+(n+1)} \\ \tilde{v}^{-(n+1)} \end{pmatrix} = \frac{1}{1+(\mu_{(j+1/2)}/2)^2} \begin{pmatrix} (1-(\mu_{(j+1/2)}/2)^2)y^{-1} & -\mu_{(j+1/2)} \\ \mu_{(j+1/2)} & (1-(\mu(j+1/2)/2)^2)y \end{pmatrix} \begin{pmatrix} \tilde{v}^{+(n)} \\ \tilde{v}^{-(n)} \end{pmatrix}, \tag{22}$$

and

$$\begin{pmatrix} \tilde{J}^{+(n+1)} \\ \tilde{J}^{-(n+1)} \end{pmatrix} = \begin{pmatrix} (1+\mu_{(j+1/2)})y^{-1} & \mu_{(j+1/2)} \\ -\mu_{(j+1/2)} & (1-\mu_{(j+1/2)})y \end{pmatrix} \begin{pmatrix} \tilde{J}^{+(n)} \\ \tilde{J}^{-(n)} \end{pmatrix}. \tag{23}$$

Local stability depends on the norm of the matrix (amplification matrix) in (22) or (23), that is, it depends upon the spectral radius of the matrix (i.e., the magnitude of its largest eigenvalue). The von Neumann condition for stability is that the eigenvalues of an amplification matrix, $\lambda$, must satisfy:

$$|\lambda| \leq 1 + \mathcal{O}(h), \quad h \to 0.$$

This condition is both sufficient and necessary only in the case the amplification matrix is normal (i.e., commutes with its adjoint), otherwise it is just necessary (Richtmeyer & Morton, 1967, pp. 68 - 73). After some algebra, we find eigenvalues for (22) and (23), respectively, satisfy:

$$\lambda^2 - 2(1-(\mu_{(j+1/2)}/2)^2)\cos(kh)\lambda + (1+(\mu_{(j+1/2)}/2)^2)^2 = 0 \tag{24}$$

and

$$\lambda^2 - 2(\cos(kh) - \imath\mu_{(j+1/2)}\sin(kh))\lambda + 1 = 0. \tag{25}$$

Since the admittance function is continuously differentiable,

$$\mu_{(j+1/2)} \leq \mathcal{O}(h).$$

With this we see that both amplification matrices satisfy the von Neumann condition.

In the case that $\mu_{(j+1/2)} = 0$ for all $j$, we would like a stronger stability, namely $|\lambda| \leq 1$, because there are no solutions of the exact differential equation that grow when area is a constant. In this case both (24) and (25) reduce to:

$$\lambda^2 - 2\cos(kh)\lambda + 1 = 0. \tag{26}$$

and hence $|\lambda| = 1$. (Not only are the systems (20) and (21) stable in this case, but they conserve the magnitude of the dependent variable. The matrices are normal and, in fact, are the identity matrix.) We are able to meet these stability conditions because our time and space intervals, $\Delta t$ and $\Delta x$, are related by $\Delta tc/\Delta x = 1$, which is a special case of the Courant condition: $\Delta tc/\Delta x \leq 1$ (Mitchell, 1969).

We now compare the dispersion relation of waves that propagate according to the finite-difference schemes (20) and (21) to dispersion relation of the waves that are an exact solution to the differential equation. The exact solution we will use is that of propagation in an exponential tube:

$$Y(x) = \exp(\alpha x)/\rho_0 c.$$

The exact volume velocity wave traveling in the positive x direction with circular frequency, $\omega$, is given as (Lighthill, 1978):

$$J^+ = J_0^+ \exp[\imath\omega t - \imath((\omega/c)^2 - (\alpha/2)^2)^{1/2}x + (\alpha/2)x]. \tag{27}$$

In terms of the dependent variable $v^+$, the solution is:

$$v^+ = v_0^+ \exp[\imath\omega t - \imath((\omega/c)^2 - (\alpha/2)^2)^{1/2}x]. \tag{28}$$

The phase, $\Theta$, is the same for both (27) and (28):

$$\Theta = \omega t - ((\omega/c)^2 - (\alpha/2)^2)^{1/2}x. \tag{29}$$

The dispersion relation is a relationship between the time and spatial dependence of the phase function. More exactly, let:

$$\omega = \frac{\partial\Theta}{\partial t}, \quad k = -\frac{\partial\Theta}{\partial x}.$$

Then the dispersion relation is of the form: $g(\omega, k) = 0$. The dispersion relation for the exact solutions is:

$$\left(\frac{\omega}{c}\right)^2 = k^2 + \left(\frac{\alpha}{2}\right)^2. \tag{30}$$

The dispersion relation for the finite-difference approximations can be derived by performing a Fourier transform in both space and time. Let:

$$v_{(j)}^{+(n)} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \tilde{\mathcal{V}}^+(\omega, k) y^j z^n \, d\omega dk,$$

where

$$z = \exp(\imath\omega h/c), \quad y = \exp(\imath kh).$$

Substituting into the finite-difference approximations (20) and (21):

$$\begin{pmatrix} \tilde{\mathcal{V}}^+ \\ \tilde{\mathcal{V}}^- \end{pmatrix} = \frac{1}{1 + (\mu_{(j+1/2)}/2)^2} \begin{pmatrix} (1 - (\mu_{(j+1/2)}/2)^2)y^{-1}z^{-1} & -\mu_{(j+1/2)}z^{-1} \\ \mu_{(j+1/2)}z^{-1} & (1 - (\mu_{(j+1/2)}/2)^2 yz^{-1} \end{pmatrix} \begin{pmatrix} \tilde{\mathcal{V}}^+ \\ \tilde{\mathcal{V}}^- \end{pmatrix} \tag{31}$$

$$\begin{pmatrix} \tilde{\mathcal{J}}^+ \\ \tilde{\mathcal{J}}^- \end{pmatrix} = \begin{pmatrix} (1 + \mu_{(j+1/2)})y^{-1}z^{-1} & \mu_{(j+1/2)}z^{-1} \\ -\mu_{(j+1/2)}z^{-1} & (1 - \mu_{(j+1/2)})yz^{-1} \end{pmatrix} \begin{pmatrix} \tilde{\mathcal{J}}^+ \\ \tilde{\mathcal{J}}^- \end{pmatrix} \tag{32}$$

The resulting systems of homogeneous linear equations must have determinant zero. Both systems satisfy:

$$\left(\frac{\omega}{c}\right)^2 = k^2 + \left(\frac{\alpha}{2}\right)^2 + \mathcal{O}\left(h^2\right), \tag{33}$$

where the neglected terms include the factors: $k^4h^2$, $(\omega/c)^4h^2$, $\alpha^4h^2$, $\alpha^3kh^2$, and $\alpha k^3h^2$. In order to keep dispersion errors small we must keep the spatial divisions small with respect to wavelength, and time divisions small with respect to wave period. Also, as before, the admittance function must be smooth: the rate of change of area with respect to $x$ must not be too large. From the above, both finite-difference schemes, (20) and (21), are seen to provide, practically, the same approximation to the dispersion relation to the original differential equation.

## Conclusion

The computational scheme resulting from the Kelly-Lochbaum reflection model has been put into the context of a finite-difference approximation to the differential equation. With a couple of minor changes, we were able to derive a finite-difference approximation with better truncation error properties, without giving anything up in terms of the von Neumann stability conditions and dispersion. We do not necessarily recommend this modified scheme for computational purposes, since the full initial-boundary value problem has not been considered.

There are many numerical methods that can be considered. One such is the Lax-Wendroff scheme, which has at least as good truncation error as the modified scheme presented here (Mitchell, 1969). Another method is integrating along characteristics in the manner of solving simultaneous ordinary differential equations, where predictor-correctors could be used (Thomas, 1954). Portnoff (1973) considered an implicit scheme for solving the differential equation. Implicit schemes are attractive because stability does not depend on small time step sizes and the boundary conditions are easily incorporated. However, there is a trade-off in terms of computational ease, where implicit schemes involve at least one matrix inversion to update all spatial positions simultaneously.

In this note, we took the starting point as a differential equation describing acoustic wave propagation which can be derived from conservation laws and under known approximations. A numerical method can be chosen for the solution of this differential equation, where bounds can be found on the error of the numerical approximation. We believe that carefully going from conservation laws to synthesis can help assess the physical model on which the synthesis is based.

## References

Forsythe, G. E. & Wasow, W. R. (1960). *Finite-difference methods for partial differential equations*. New York: John Wiley and Sons.

Higdon, R. L. (1986). Initial-boundary value problems for linear hyperbolic systems. *Society for Industrial and Applied Mathematics Review, 28*, 177-217.

Lesser, M. B., & Lewis, J. A. (1972). Applications of matched asymptotic expansion methods to acoustics. I. The Webster horn equation and the stepped duct. *Journal of the Acoustical Society of America, 51*, 1664.

Lighthill, M. J. (1978). *Waves in fluids*. Cambridge, England: Cambridge University Press.

Liljencrants, J. (1985). *Speech synthesis with a reflection-type line analog*. Unpublished doctoral dissertation, Royal Institute of Technology, Stockholm, Sweden.

Markel, J. D., & Gray, A. H. (1976). *Linear prediction of speech*. Berlin: Springer-Verlag.

Maeda, S. (1982). A digital simulation method of the vocal-tract system. *Speech Communication, 1*, 99-229.

Mitchell, A. R. (1969). *Computational methods in partial differential equations*. New York: John Wiley and Sons.

Portnoff, M. R. (1973). *A quasi-one-dimensional digital simulation for the time-varying vocal tract*. Unpublished master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Richtmeyer, R. D., & Morton, K. W. (1967). *Difference methods for initial value problems*. New York: John Wiley and Sons.

Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for speech research. *Journal of the Acoustical Society of America, 70*, 321 - 328.

Thomas, L. H. (1954). Computation of one dimensional flows including shocks. *Communications in Pure and Applied Mathematics*, 7, 195 - 206.

Thompson, C. (1984). Linear inviscid wave propogation in a waveguide having a single boundary discontinuity: Part I: Theory. *Journal of the Acoustical Society of America, 75*, 346-355.

Trefethen, L. N. (1982). Group velocity in finite difference schemes. *Society for Industrial and Applied Mathematics Review 24*, 113 - 136.

## Appendix

To derive equation (7) from equation (4), left multiply equation (4) by **GB**:

$$\mathbf{I}\frac{\partial}{\partial t}(\mathbf{GB})\mathbf{U} + c(\mathbf{GB})\mathbf{A}(\mathbf{GB})^{-1}(\mathbf{GB})\frac{\partial}{\partial \mathbf{x}}\mathbf{U} = \mathbf{0}.$$

Using the definitions of **V** and **H**, and using the product rule for differentiation:

$$\mathbf{I}\frac{\partial}{\partial t}\mathbf{V} + c\mathbf{H}\frac{\partial}{\partial \mathbf{x}}\mathbf{V} - c\mathbf{H}(\frac{\partial \mathbf{GB}}{\partial \mathbf{x}})(\mathbf{GB})^{-1}\mathbf{V} = \mathbf{0}.$$

Equation (7) results if:

$$\mathbf{K} = \mathbf{H}(\frac{\partial \mathbf{GB}}{\partial \mathbf{x}})(\mathbf{GB})^{-1}.$$

# TYPE AND NUMBER OF VIOLATIONS AND THE GRAMMATICAL CONGRUENCY EFFECT IN LEXICAL DECISION*

**G. Lukatela,[**] A. Kostić,[**] D. Todorović,[**] C. Carello,[†] and M. T. Turvey[‡]**

**Abstract.** *An experiment was conducted in the Serbo-Croatian language in which native speakers/readers made lexical decisions on inflected nouns and legally inflected pseudonouns following inflected possessive pronouns. A possessive pronoun and the noun or pseudonoun that followed it could agree in case, gender, and number (0 violations), disagree in either case or gender or number (1 violation) or disagree simultaneously on two of the three (2 violations). A grammatical congruency effect was observed for both nouns and pseudonouns. Acceptance latencies were shorter and rejection latencies were longer for inflectional agreement than inflectional disagreement. However, for neither nouns nor pseudonouns was the magnitude of the effect influenced by the type or number of violations. The results are discussed in terms of (1) the automaticity of syntactic processes and (2) the properties of a decision making device (specially tailored to rapid lexical evaluations) relative to the properties of the language processor.*

## INTRODUCTION

A growing body of evidence supports the notion that syntactical or grammatical relatedness colors the way in which one word affects the processing of another. Investigations with English language materials address this issue by violating the natural ordering of parts of speech. For example, lexical decision to a target is speeded when the context-target pair is ordered legally relative to when it is ordered illegally (e.g., men-swear vs. whose-swear [Goodman, McClelland, & Gibbs, 1981]; "For now the happy family lives with BATTERIES" vs. "For now the happy family lives with FORMULATE" [Wright & Garrett, 1984]). In contrast, investigations with Serbo-Croatian materials have been able to preserve the ordinary adjacencies of parts of speech because grammatical violations can be introduced at the level of inflected morphemes. Grammatically acceptable pronoun-verb pairs must agree in person and number while adjective-noun pairs must

---

[*] *Psychological Research, 49*(1), 37-43

[**] University of Belgrade

[†] State University of New York at Binghamton

[‡] Also University of Connecticut

*Acknowledgment.* This research was supported in part by NICHD Grants HD-08495 and HD-01994 to the University of Belgrade and Haskins Laboratories, respectively.

agree in case, number, and gender. Violations of these relationships result in a grammatical congruency effect, viz., lexical decision to targets in a grammatically incongruent context are slow relative to those same targets in grammatically congruent contexts. As examples, lexical decisions to verb targets are faster when the preceding personal pronoun agrees in person than when it does not (Lukatela, Moraca, Stojnov, Savić, Katz, & Turvey, 1982); decision times to nouns with a case inflection appropriate for a preceding preposition are speeded relative to those with an inappropriate inflection (Lukatela, Kostić, Feldman, & Turvey, 1983); slowed decision times are found for violations of case agreement between adjectives and nouns or legally inflected pseudoadjectives and nouns (Gurjanov, Lukatela, Moskovljević, Savić, & Turvey, 1985); and nouns that agree with their possessive pronoun contexts in gender are lexically evaluated faster than those that do not agree (Gurjanov, G. Lukatela, K. Lukatela, Savić, & Turvey, 1986).

It has been argued that syntactic influences on lexical decision are post-lexical (Gurjanov et al., 1985, 1986; Seidenberg, Waters, Sanders, & Langer, 1984; West & Stanovich, 1982). That is to say, unlike the spreading activation among particular lexical items that is conjectured for associative priming (deGroot, 1983), the grammatical congruency effect is thought to be the result of a check on grammatical coherency of the given context-target pair (cf. deGroot, Thomassen, & Hudson, 1982; Gurjanov et al., 1986). The reason is quite simple: If the congruency effect were the result of spreading activation, then a prime would have to activate all words of a given type (e.g., all nouns of a particular case). It seems unlikely, therefore, that relations among lexical entries are responsible for syntactical influences on lexical decision.

Let us, then, provide a framework for this coherence checker. The central notion is that the language processor is composed of three relatively autonomous devices. One accesses lexical representations of each member of an arrangement of words, another assigns a syntactical structure to the arrangement of words, and the third assigns meaning to the arrangement of words (cf. Forster, 1979). In the course of normal language comprehension, all three devices are necessary. In the experimentally contrived situation of a lexical decision task, although it would seem that the lexical processor is all that is required, the other devices cannot be disengaged. With a grammatically congruent context-target pair, all devices provide positive output (i.e., each performs its usual function) so that the job of the decision-making mechanism is easy. With a grammatically incongruent pair, however, the syntactic processor balks because part of the information made available by the lexical processor is that, for example, the context is masculine and the target is feminine. The lexical decision mechanism must overcome the negative bias from the syntactical processor (cf. deGroot, 1985; West & Stanovich, 1982), resulting in slower decision times.

It was mentioned earlier that grammatical congruency in the Serbo-Croatian language is defined over several dimensions. At issue in the present investigation is whether or not the congruency effect for possessive pronoun-noun pairs is influenced by: (1) which grammatical dimension—gender, number, or case—is violated, or (2) how many grammatical dimensions are violated. In other words, is the negative bias that is induced by the coherence check altered by the type or the extent of grammatical violation?

This question is directed primarily at the nature of the device that makes a decision about a word's lexical status on the basis of the information it receives from the largely independent lexical, syntactic, and message processors. These latter processes are presumed to be "hard molded, hard algorithmed." The decision making device, on the other hand, is presumed to be "soft molded, soft algorithmed." It represents the fact that an ordinary speaker/reader of the language has temporarily made him or herself into a special purpose mechanism—one geared to reporting rapidly on the lexical status of printed letter strings. One could imagine that it is in the nature of this soft-molded decision making device to weight the outcomes of the lexical, syntactic, and message processors. In a lexical decision experiment, for example, the lexical processor ought to be weighted most heavily. The value of the message processor would depend on how informative it is, given the constraints of the experimental situation. To anticipate our method, the present investigation simply uses some form of the possessive pronouns MY or YOUR on every trial. The message processor, therefore, is relatively noninformative and ought to be weighted accordingly. In contrast, numerous investigations of the effect of minimal grammatical contexts—for example, a single, closed class word with an inflection appropriate or inappropriate for the target—reveals that considerable weight is given to the syntactic processor in lexical decision.

Obviously, the more that the outcomes of the three processors concur, the larger the probability that the lexical decision device will succeed in making a decision in a determined period of time. However, before a soft molded decision device can operate on, say, grammatical incongruency, it must receive information that incongruency of some type has been detected. This information must come from the hard molded syntactic processor. It is reasonable, therefore, to expect the soft molded decision maker to be sensitive to the speed of detection of an incongruity. One could hypothesize that the speed of detection might depend on the type and/or number of grammatical violations (case violation might be considered more egregious than—and be detected faster than—gender violation; two violations of any type might be detected faster than any single one; and so on). In experimental terms, these hypothesized properties of the decision making device would be realized as lexical decision times on nouns in the context of possessive adjectives that (1) differ significantly as a function of the type of incongruency and (2) increase as a direct function of the number of incongruencies.

If the outcome of the experiment runs counter to the outlined hypotheses and shows no differential effect as a function of type or number of violations, then this lack of an effect can just as plausibly be ascribed to the real structural—i.e., hard molded—processor as to the decision maker. A little thought suggests that in order to do its real world job effectively, the hard molded syntactic processor might only need to detect the fact that there is or is not a grammatical incongruency. Therefore, a self-terminating scan of grammatical features that is associated with binary coherence checks seems to be a plausible model of the syntactic processor. In experimental terms, this latter perspective on the decision making device suggests that the lexical decision times for any type and any number of incongruencies will be the same and that they will be significantly slower than zero incongruencies.

The present experiment addresses these experimental predictions by observing the effects of different grammatical relations (1) between possessive pronouns (sometimes referred to as possessive adjectives) and nouns and (2) between possessive pronouns and pseudonouns. Pseudonouns are created from real nouns by substituting for one of the letters in the stem. Their inflected endings, therefore, are legal noun endings. In consequence, grammatical congruency can be defined between a possessive pronoun and a pseudonoun in the same way that it can be defined between a possessive pronoun and a noun. To the extent that grammatical relations are sustained purely by inflectional morphemes,[1] equivalent effects should be observed for acceptance latencies (nouns) and rejection latencies (pseudonouns). In order to avoid a confound between grammatical and physical congruence of inflectional endings, targets were limited to feminine singular nouns in the dative case. The inflectional endings of such nouns (-I) and their congruent possessive adjectives (MOJOJ and TVOJOJ) are physically dissimilar.

The aforementioned equivalence between effects obtained with acceptance and rejection latencies has been noted in two previous grammatical congruency experiments (Lukatela et al., 1982, 1983). The data from a study that used possessive pronoun-noun pairs (Gurjanov et al., 1986), as the present experiment does, were ambiguous about the equivalence.

## Method

### Subjects

Seventy-two students from the Department of Psychology in the Faculty of Philosophy at the University of Belgrade participated in the experiment in partial fulfillment of a course requirement. All subjects had previously participated in reaction time experiments.

### Materials

Targets were selected from a basic set of 80 nouns, all of the CCVCV type (e.g., PTICA, "bird") drawn from the mid-frequency range (Dj. Kostić, 1965). Corresponding pseudonouns were formed using an entirely different set of 80 comparable nouns and changing one letter in the stem of each (leaving the inflectional morpheme intact). Of the 160 context-target pairs (see Appendix), 100 were test trials and 60 were filler trials included to equate the number of congruent and incongruent pairs seen by a given subject. The fillers were not analyzed.

All targets in the test trials were singular feminine nouns of Class A (after Bidwell, 1970) in the dative case (where the ending is /i/). Fifty of these were paired with possessive pronouns (half first person [MY] and half second person [YOUR]) to generate five types of situations containing ten tokens of each type: one set with no violations, three sets with one violation (where case was accusative, gender was masculine, or number was plural) and one set with two violations (where gender was masculine and, simultaneously number was plural). Fifty corresponding context-pseudonoun pairs were similarly constructed. In addition to precluding physical similarity of inflectional endings for contexts and targets, the selection restrictions ensured that only unique violation types were produced (test trials included only Class A feminine singular nouns in the dative case, case violations were introduced solely with accusative contexts, and the two-violation condition was limited to gender + number). (For example, Type A feminine nominative singular and Type O masculine genitive singular both end in /a/ so that, had such targets been used, the extent of the violation would be ambiguous.)

---

[1] Although it is assumed that pseudowords have no lexical entry, there is evidence that some pseudowords derived from real words may access the lexical entry of the source words (e.g., Martin, 1982; but see Chambers, 1979). Of course, this would affect syntactically congruent and incongruent situations to the same extent.

For the filler trials, 10 feminine singular accusative, 10 masculine singular dative, and 10 feminine plural dative nouns were paired with appropriate pronouns, as were a corresponding set of pseudonouns.

### Design

Each subject saw 80 pronoun-noun and 80 pronoun-pseudonoun pairs, half of which were grammatically congruent and half of which contained at least one violation. Of the incongruent pairs, there were equal numbers of case, gender, number, and gender-plus-number violations. A given subject never encountered a given target more than once.

### Procedure

A subject was seated before the CRT of an Apple IIe computer in a dimly lit room. A fixation point was centered on the screen. On each trial, the subject heard a brief warning signal after which a possessive pronoun appeared for 300 ms centered above the fixation point. After a 300 ms interstimulus interval a noun or pseudonoun appeared below the fixation point for 1400 ms. All letter strings appeared in uppercase Roman. Subjects were instructed to decide as rapidly as possible whether or not the second letter string was a word. To ensure that subjects were reading the contexts, they were occasionally asked to report both stimuli after the lexical decision had been made. Decisions were indicated by depressing a telegraph key with both thumbs for a "No" response or by depressing a slightly further key with both forefingers for a "Yes" response. Latencies were measured from the onset of the target. If the response latency was longer than 1400 ms, a message appeared on the screen requesting that the subject respond more quickly. The experimental sequence was preceded by a practice sequence of 20 different context-target pairs.

### Results and Discussion

Latencies in excess of 1400 ms and less than 400 ms were excluded from the analysis. The means of the subjects' latencies and errors for the three types of violations with noun and pseudonoun targets are presented in Table 1. Inspection of Table 1 suggests that for single violations, decision latencies were not distinguished by type of violation. For the noun latencies and errors the $F$ ratios were less than unity by both the subjects and stimuli analyses. The $F$ by the subjects' analysis for the pseudoword latencies exceeded unity but was not significant, $F(2,142) = 1.63, MSe = 1288, p > .10$. The three other $F$ tests on the pseudoword data (latencies by stimuli and errors by subjects and stimuli) yielded values less than unity. In short, type of violation did not differentially affect word and pseudoword latencies and errors.

Given this fact, the latency and error data were collapsed over the type variable to yield three sets of means corresponding to 0, 1, and 2 grammatical violations and these are presented in Table 2. The effect of number of violations was evaluated on these means. Noun latencies were significantly affected by number according to both the subjects and the stimuli analyses, $F(2,142) = 5.36, MSe = 1402, p < .01$ and $F(2,118) = 5.95, MSe = 1311, p < .01$, respectively. The same statistical outcomes were obtained for the pseudonoun latencies: $F(2,142) =$

### Table 1

### Lexical Decision as a Function of Type of Grammatical Violation

| | Type of Violation | | |
|---|---|---|---|
| Target | Case | Gender | Number |
| Noun | 671[a] | 675 | 671 |
| | 4.4[b] | 6.0 | 6.0 |
| Pseudonoun | 718 | 708 | 717 |
| | 2.6 | 3.5 | 2.1 |

[a] latency (ms)
[b] error (percent)

### Table 2

### Lexical Decision as a Function of Number of Violations

| | Number of Violations | | |
|---|---|---|---|
| Target | 0 | 1 | 2 |
| Noun | 656[a] | 672 | 675 |
| | 3.2[b] | 5.5 | 6.1 |
| Pseudonoun | 730 | 714 | 715 |
| | 3.3 | 2.7 | 3.6 |

[a] latency (ms)
[b] error (percent)

$4.65, MSe = 1147, p < .01$ by the subjects analysis and $F(2,118) = 4.86, p < .01$ by the stimuli analysis. Errors in noun decision making were significantly affected by number of violations according to both the subjects and stimuli analyses: $F(2,142) = 4.97, MSe = 34, p < .01$ and $F(2,118) = 7.37, Mse = 31, p < .001$. In contrast, number did not affect pseudonoun errors. The ANOVA on subjects and stimuli means both yielded $F$ ratios less than unity.

Protected t-tests (where the error term from the ANOVA is used as the estimate of the variance; see Cohen & Cohen, 1975) were conducted on the means for the 1 versus 2 violations. No significant differences were obtained.

The results of the experiment are fairly straightforward. First, there was a grammatical congruency effect, and it was observed for both nouns and pseudonouns. Second, the magnitude of the effect for both nouns and pseudonouns was indifferent to the type and the number of grammatical violations.

Let us consider the first result. Possessive pronoun-noun pairings that were in full grammatical agreement were associated with faster lexical decisions than possessive pronoun-noun pairings that disagreed on one or two grammatical dimensions. Similarly, possessive pronoun-pseudonoun pairings that were in full grammatical agreement (the pseudonoun's inflection agreed in case, gender, and number with the possessive pronoun's inflection) were associated with slower rejection latencies than pairings in which the agreement was incomplete by one or two dimensions. The magnitude of the grammatical congruency effect in the noun latency data was 16 ms for zero versus one violation and 19 ms for zero versus two violations. Gurjanov et al. (1986) obtained a congruency effect for zero versus one violation of the order of 51 ms (calculated from the data on feminine nouns preceded by possessive pronouns reported in their Table 2). In the course of the latter experiment, only one type of disagreement ever occurred, namely, in gender. It contrasts, therefore, with the present experiment in which all three types of possible disagreement occurred and in which the number of disagreements was frequently two. The large difference in the magnitudes of the congruency effect defined over possessive pronoun-noun pairs in the two experiments is probably attributable to these differences in the homogeneity of grammatical manipulations. The situation may be analogous to that in associative priming experiments. Tweedy, Lapinski, and Schvaneveldt (1977) showed that the facilitation due to an associative context was greater with a larger proportion of associative trials. They interpreted this result within Posner and Snyder's (1975) two-factor theory of attention. Focusing on the conscious attentional component, Tweedy et al. (1977) argued that the subjects' expectation concerning the relatedness of the items allows for a specialized post-lexical control strategy (cf. Shiffrin & Schneider, 1977) to be brought into effect. In principle, the decision making device in the Gurjanov et al. (1986) experiments could concentrate on just the gender dimension. The concentration in the present experiment could not have been as focused because the subjects' expectancies were that any one of the dimensions of grammatical agreement could be violated with near equal probability.

The magnitude of the grammatical congruency effect on word (noun) latencies in the present experiment compares favorably with the magnitudes of syntactical congruency effects reported for English language two-word sequences by Goodman et al. (1981) and Seidenberg et al. (1984). In the two experiments of Goodman et al. the magnitudes were 19 ms and 15 ms. In the single experiment of Seidenberg et al. the magnitude was 13 ms. A further favorable comparison is to be found between the respective error productions. In the present experiment the percent error for the congruent condition was 3.19. For the single and double incongruency conditions the percent errors were 5.42 and 6.11, respectively, to yield congruent-incongruent differences of −2.24 percent and −2.92 percent. Significant differences in error production between congruent and incongruent conditions on the order of −4.0 percent and −1.3 percent were reported respectively, for the first of Goodman et al.'s experiments and for the Seidenberg et al. experiment. In the Gurjanov et al. (1986) study, the congruency-incongruency error production difference (averaged over masculine and feminine nouns of typical and atypical declension) amounted to −2.7 percent.

The grammatical congruency effect in the pseudonoun latency data was −16 ms for the 0 versus 1 comparison and −15 ms for the 0 versus 2 comparison. These rejection latency differences complement the acceptance latency differences and they concur in this respect with the results of several previous experiments that used pseudoverbs and pseudoadjectives as well as pseudonouns. We will summarize these findings briefly before elaborating the significance of grammatical effects with pseudowords.

The preposition-noun experiment of Lukatela et al. (1982) included pseudonouns that were mostly but not exclusively generated by the substitution of the first letter of a noun keeping the inflected ending legal. An interaction between preposition and pseudonoun type (nominative-like, dative/locative-like, instrumental-like) was obtained with subject variability as the error term but not with stimulus variability as the error term. The data suggested that where the inflection of a pseudonoun agreed with the preceding preposition, the rejection latencies were slowed (by approximately 18 ms) relative to when they were in disagreement. For the noun targets grammatical agreement with the preceding preposition hastened (by approximately 28 ms) positive decisions relative to grammatical disagreement. In the pronoun-verb experiments of Lukatela et al. (1983) all pseudoverb stimuli were inflected with verb endings. They were created by single letter substitution in the stems of the verbs. These experiments also provided evidence for complementary effects between the positive and negative latencies. Taking the first experiment of Lukatela et al. (1982) as an example, grammatical congruency resulted in faster (by 128 ms) positive decisions and slower (by 27 ms) negative decisions. Finally, the experiments of Gurjanov et al. (1985) that examined adjective-noun pairings should be mentioned. These experiments found no evidence of a grammatical congruency effect with pseudonoun targets. They did demonstrate, however, a grammatical congruency effect with *pseudoadjective*-noun pairs (that is, on positive decision latencies) that was as large as the effect for adjective-noun pairs.

The significance of demonstrating grammatical congruency effects with legally inflected pseudowords as either contexts or targets is that it points to the main carriers of grammatical information, the inflectional morphemes, as largely responsible for the effect. In more theoretical terms, it lends support to the hypothesis that the syntactic level of processing operates relatively independently from the semantic-interpretative processes (Forster's message processor). When pseudowords are used as either contexts or targets, the "word" sequence is meaningless. Consequently, one cannot appeal to a process of sentence comprehension to effect, in top-down fashion, the syntactic analysis. Further, when pseudowords are used as either contexts or targets, the lexical processor must deliver definitional information, to use Fodor's (1983) term, pertaining to the grammatical function of the pseudoword's inflection. The implication is that lexical processes work with a morphemic inventory and can effectively distinguish morphemic constituents in the absence of activating full (that is, word) lexical entries. That the grammatical congruency effect is demonstrable with pseudowords means that the lexical processes provide acceptable inputs to the syntactic processes. We must, nevertheless, be careful of carrying this line of argument too far. The grammatical congruency effect is less reliable for pseudowords than words. And this difference is probably telling us (not surprisingly) that the stem as well as the suffix is a source of grammatical information. The lexical processor working with words rather than the constituents of words can more reliably furnish definitional information about the parts of speech. Serbo-Croatian nouns share many of their inflections with other word types (most notably with adjectives but also with the cardinal numerals). To the extent that stem information is not accessed, the identity of a letter string as a noun is less clear and the lexical processor is less able to provide acceptable resources for the syntactic operations.

Another reason that the grammatical congruency effect is more difficult to reveal with pseudoword targets is that the process of isolating affixal information in pseudowords may be slower than in words. In consequence, the lexical search determining the absence of a pseudoword's entry may often be completed before affixal information about the pseudoword has been discerned (Wright & Garrett, 1984). Under these conditions no contribution of the syntactic processor would be expected.

The second result of the present experiment was that the magnitude of the grammatical congruency effect, for both nouns and pseudonouns, was indifferent to the type of violation and to the number of violations (one or two). In terms of the arguments raised in the introduction, this result suggests that the information of relevance to the decision making process is merely that the two words do not agree grammatically. Type of disagreement and the number of disagreements do not affect the magnitude of the negative bias (that hinders positive decisions and aids negative decisions). Each type of grammatical disagreement (case, gender, and number) contrasted with complete agreement. This fact of a grammatical congruency effect defined with respect to each violation suggests that, in the experiment, syntactic processors were evaluating all three grammatical relations between the possessive pronoun context and the noun or pseudonoun target. From the perspective of the job that these processors ordinarily perform in everyday sentence comprehension, namely, assigning grammatical structure to word sequences, it may well be that the assignment relies differentially on case, gender, and number information. This possibility cannot be ruled out by the fact that in the present experiment each type of grammatical disagreement contrasted with full agreement to the same degree and by the related fact that two disagreements were no worse than one.

Inferences from lexical decision data to underlying linguistic mechanisms have to contend with the soft algorithmical capabilities assembled specifically for the task. As suggested in the introduction, it is useful to construe a subject in a lexical decision task as assembling him or herself into a device specially tailored to the goal of passing rapid judgment on the lexical status of a letter string. The subject, of course, is a language processor—a complex device that ordinarily analyzes multiple embeddings of linguistic structures of different grains, and does so on line. Fashioning a device tailored to lexical decision can be regarded as the fashioning of an *alternative description* of the language processor (see Pattee, 1972, for a general argument of this kind with regard to biological functions). This alternative (simpler) description makes explicit some of the detailed processing that is implicit in ordinary sentence comprehension. The important point to be underscored is that the special purpose lexical decision device as an alternative (simpler) description of the language processor is selective. It does not make explicit all of the processing detail. Thus, it suffices for lexical decision to make explicit grammatical conformity. The nature and time course of the processing details that determine grammatical conformity remain, however, largely implicit.

## References

Bidwell, C. E. (1970). Serbo-Croatian nominal inflection. In P. Filopovic (Ed.), *The Yugoslav Serbo-Croatian English contrastive project*. Washington, D.C.: Center for Applied Linguistics.

Chambers, S. M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior, 18*, 225-241.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

deGroot, A. M. B. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior, 22*, 417-436.

deGroot, A. M. B. (1985). Word-context effects in word naming and lexical decision. *Quarterly Journal of Experimental Psychology, 37A*, 281-297.

deGroot, A. M. B., Thomassen, A. J. W. M., & Hudson, P. T. (1982). Associative facilitation of word recognition as measured from a neutral prime. *Memory & Cognition, 10*, 358-370.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. Walker (Eds.), *Sentence processing*. Hillsdale, NJ: Erlbaum.

Goodman, G. C., McClelland, J. L., & Gibbs, R. W. (1981). The role of syntactic context in word recognition. *Memory & Cognition, 9*, 580-586.

Gurjanov, M., Lukatela, G., Lukatela, K., Savić, M., & Turvey, M. T. (1986). Grammatical priming of inflected nouns by the gender of possessive adjectives. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11*, 692-701.

Gurjanov, M., Lukatela, G., Moskovljević, J., Savić, M., & Turvey, M. T. (1985). Grammatical priming of inflected nouns by inflected adjectives. *Cognition, 19*, 55-71.

Kostić, Dj. (1965). *Frequency of occurrence of nouns in Serbo-Croatian*. Unpublished manuscript, Institute of Experimental Phonetics and Speech Pathology, University of Belgrade.

Lukatela, G., Kostić, A., Feldman, L. B., & Turvey, M. T. (1983). Grammatical priming of inflected nouns. *Memory & Cognition, 11*, 59-63.

Lukatela, G., Moraca, J., Stojnov, D., Savić, M., Katz, L., & Turvey, M. T. (1982). Grammatical priming effects between pronouns and inflected verb forms. *Psychological Research (Psychologische Forschung), 44*, 297-311.

Martin, R. C. (1982). The pseudohomophone effect: The role of visual similarity in non-word decisions. *Quarterly Journal of Experimental Psychology, 34A*, 395-490.

Pattee, H. H. (1972). Laws and constraints, symbols and language. In C. H. Waddington (Ed.), *Towards a theoretical biology*. Chicago: Atherton-Aldine.

Posner, M. I., & Snyder, C. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V*. London: Academic Press.

Schiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing II. Perceptual learning, automatic attending and a general theory. *Psychological Research, 84*, 127-190.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word recognition. *Memory & Cognition, 12*, 315-328.

Tweedy, J. R., Lapinski, R. H., & Schvaneveldt, R. W. (1977). Semantic context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory & Cognition, 5*, 84-89.

West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 5*, 385-399.

Wright, B., & Garrett, M. (1984). Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition, 12*, 31-45.

## APPENDIX

For the experimental situations, all word (•) and pseudoword targets are feminine singular nouns in the dative case. Possessive adjective contexts are either grammatically congruent or violate case, number, gender, or number + gender.

| Context | Target | Context | Target | Context | Target |
|---|---|---|---|---|---|
| TVOJ | PLOŽI | MOJOJ | DURKI | TVOJU | STURI |
| MOJIM | •FRULI | MOM | ROČKI | MOJ | SKEPI |
| TVOJU | •VRANI | TVOM | SKABI | TVOJOJ | •KREDI |
| MOJ | ČERPI | TVOJIM | •SNAJI | TVOJ | VRSTI |
| MOJU | ŠPALI | MOJOJ | BLUCI | TVOJU | DRIGI |
| TVOJOJ | RESMI | MOM | •GRIVI | MOJ | •KLICI |
| MOJOJ | •PLIMI | TVOJIM | SRIZI | TVOM | PRUPI |
| MOJIM | LUSNI | TVOM | •TRAVI | MOJ | •BRAVI |
| MOJU | •GRUPI | MOM | •FLOTI | MOJOJ | GRAVI |
| MOJIM | •SKELI | MOJIM | TRAPI | TVOJ | •OLUJI |
| MOJU | DRANI | TVOM | BETKI | TVOJU | •ŠTALI |
| TVOJ | TRABI | TVOJOJ | OBRVI | MOJU | •PTICI |
| MOJ | TRANI | TVOJU | FLUDI | MOJIM | •ČETKI |
| MOJU | DOLKI | TVOJIM | ŠTUKI | MOJOJ | •OBALI |
| MOM | GRODI | TVOJU | •ZVEZDI | MOJOJ | •ZEMLJI |
| MOJ | DITRI | TVOJIM | •KRIZI | TVOJ | •SKALI |
| MOM | •MREŽI | TVOJIM | •VATRI | | |
| TVOJOJ | TRIVI | MOJU | BRULI | MOJOJ | •BREZI |
| MOJIM | KLEDI | MOJU | •STAZI | MOJOJ | KIRTI |
| TVOM | •BRADI | MOM | •PESMI | TVOJOJ | DIBRI |
| TVOJIM | SLUCI | TVOJ | •TABLI | TVOM | TASVI |
| MOJIM | KARČI | TVOJ | MASLI | TVOJOJ | •TETKI |
| MOM | STEJI | TVOJU | •BREZI | TVOJOJ | TLASI |
| MOJIM | •PRIČI | MOM | •KRAVI | MOJU | KROBI |
| TVOJIM | •STENI | MOJOJ | •KLUPI | MOJ | VREKI |
| TVOM | TRUKI | TVOJ | ROSTI | TVOJIM | •BLUZI |
| MOJIM | GRACI | MOJ | •FRULI | TVOJOJ | •TRUPI |
| MOM | •JAKNI | TVOJU | BLAVI | TVOJOJ | PLIDI |
| MOJ | •POŠTI | MOJOJ | •SESTRI | TVOJIM | KRESI |
| TVOM | DASPI | MOJU | •GLAVI | MOJIM | GLUFTI |
| TVOM | •GREDI | TVOJ | •CIPELI | MOJU | •PLOČI |
| MOM | •PLAŽI | TVOJOJ | PIGLI | MOJ | KORVI |
| TVOJ | •SVILI | TVOJU | TREZI | TVOM | PUSNI |
| TVOJIM | PALKI | TVOJU | •SARMI | | |

For the filler situations, nouns and pseudonouns were either feminine singular accusative, masculine singular dative, or feminine plural dative.

| | |
|---|---|
| TVOJU | •BARKU |
| TVOJIM | •MAJKAMA |
| MOM | •BRATU |
| MOM | •DRUGU |
| MOJIM | NORAMA |
| MOJIM | •PROBAMA |
| TVOM | •STRICU |
| TVOM | KROBU |
| MOJIM | NARAMA |
| MOM | •LAKTU |
| TVOJIM | •LAMPAMA |
| MOJU | •FARSU |
| TVOJU | MATLU |
| MOM | NORKU |
| TVOJU | •DRAMU |
| TVOJIM | •NOČIMA |
| MOJU | ČOTKU |
| MOJIM | •UŠIMA |
| MOJIM | •BUNDAMA |
| TVOJIM | PURKAMA |
| MOJIM | •RUKAMA |
| MOJU | •OLOVKU |
| TVOJIM | •NOGAMA |
| TVOJU | TARTU |
| MOJU | LOPKU |
| TVOJIM | PLORAMA |
| MOJU | •BORBU |
| MOJU | PAMKU |
| MOJU | •POLKU |
| MOM | •DEČKU |
| MOJU | GURBU |
| MOM | TARKU |
| MOJIM | •SLIKAMA |
| MOM | SPANU |
| TVOJIM | •OČIMA |
| MOM | •VOLANU |
| MOJIM | DAKAMA |
| TVOJIM | KORKAMA |
| TVOM | KOTKU |
| MOJIM | MALFIMA |
| TVOM | •STOLU |
| MOJIM | RAMIMA |
| MOJU | BAŠNU |
| MOJU | •MAŠNU |
| MOM | SAČKU |
| MOM | •MATKU |
| MOM | GROKU |
| MOJU | KRAMU |
| MOM | •DLANU |
| TVOM | RAČKU |
| TVOM | •PRSTU |
| MOJIM | DURIMA |
| TVOJU | •DLAKU |
| MOJU | •BRIGU |
| MOJU | BASTU |
| MOM | •MOSTU |
| MOJU | GUSPU |
| MOJU | •BRANU |
| MOM | DUZRU |

# LOW CONSTRAINT FACILITATION IN LEXICAL DECISION WITH SINGLE WORD CONTEXTS*

**G. Lukatela,** Claudia Carello,† A. Kostić,** and M. T. Turvey‡**

**Abstract.** *Single word, low constraint adjective contexts were used to "prime" lexical decision to noun targets in Serbo-Croat. Semantically congruent situations consisted of adjective-noun pairs that were not highly predictable but were nonetheless plausible (e.g., GOOD-AUNT). Semantically congruent situations used pairs that were implausible (e.g., SLOW-COAT). All adjective-noun pairs were grammatically congruent and were compared to a neutral xxx baseline. In Experiment 1, at a stimulus onset asynchrony of 300 ms, congruous situations showed 59 ms of facilitation while incongruous situations did not differ from the baseline. The same pattern was repeated in Experiment 2, at a stimulus onset asynchrony of 800 ms. Congruous situations were facilitated 67 ms. Results were discussed in terms of a message-level coherence check in Forster's (1979) model of autonomous levels of language processing.*

## Introduction

The existence of facilitating sentence context effects has been considered to be of much theoretical significance. Recent interest has centered on the difference between low constraint or unfocused contexts—those for which many completions are appropriate but no one is particularly predictable—and high constraint or focused contexts—those for which a particular completion is highly predictable. The issue concerns whether or not low constraint context effects occur and, if they do, whether they can or should be interpreted as arising from generalized priming. A generalized priming interpretation means that a very large set of lexical items is primed, or the features generated are few and general, or subjects' attention is focused on a wide range of completions. Such explanations suggest that higher level knowledge and expectations can relate interactively with lower level processes such as word recognition (e.g., Sanocki, Goldman, Waltz, Cook, Epstein & Oden, 1985; Schwanenflugel & Shoben, 1985).

---

[*] *American Journal of Psychology*, in press

[**] University of Belgrade, Serbia, Yugoslavia

[†] Trinity College

[‡] Also University of Connecticut

*Acknowledgment*. This research was supported in part by National Institute of Child Health and Human Development Grants HD-08495 and HD-01994 to the University of Belgrade and Haskins Laboratories, respectively.

Such an account contrasts with approaches that maintain the autonomy of different levels of processing (e.g., Forster, 1979, 1981; West & Stanovich, 1982). The levels are separate and hierarchically arranged: The lexical processor receives input only from feature analysis; the syntactic processor receives input only from the lexical processor; the message processor receives input only from the lexical processor and the syntactic processor. Clearly, sentence context effects cannot influence lexical processing.

> ...[E]ffects due to *lexical* context (i.e., single word contexts) are entirely acceptable within this theory, since they can be described as *within* level effects rather than between level effects. That is, the lexical context effect is assumed to be mediated by structural properties entirely internal to the lexical processor itself, and no other level of processing need be involved (Forster, 1979). Viewed from this perspective, then, the possibility that lexical and sentence context effects might have different properties takes on considerable significance (Forster, 1981, p. 471).

The data from semantic sentence context effects reveal that appropriate semantic completions are fast relative to inappropriate completions. But when compared to a neutral baseline, results are mixed. For high constraint sentences, appropriate completions are usually facilitated and inappropriate completions are inhibited (Forster, 1981; Schwanenflugel & Shoben, 1985; although see Fishler & Bloom, 1979, for predictable completions that did not differ significantly from the baseline). For low constraint sentences, inappropriate completions are inhibited but appropriate completions either show no difference relative to a neutral baseline (Fischler & Bloom, 1979; Forster, 1981) or show significant facilitation that is less than that found for predictable completions (Schwanenflugel & Shoben, 1985).

The Serbo-Croatian language has provided a convenient medium for exploring low constraint contexts. Although the investigations have used syntactic rather than semantic contexts, they are nonetheless instructive for present purposes. As an inflected language, Serbo-Croat permits the creation of highly salient grammatical contexts with a single word (note, in contrast with Forster, 1981, that single words need not be simply lexical contexts). Furthermore, it does not require that word class be violated in order to obtain grammatical incongruency as is typically done with English language materials (e.g., Wright & Garrett, 1984). For example, adjectives and nouns must agree in gender (masculine, feminine, or neuter), case (e.g., nominative, dative, accusative, etc.), and number (singular or plural). When a context and target agree on these dimensions, lexical decision is faster than when one or more of the dimensions is incongruent (Gurjanov, Lukatela, Moskovljević, Savić, & Turvey, 1985). Similar effects have been found for pronoun-verb pairs with respect to person (Lukatela, Moraca, Stojnov, Savić, Katz, & Turvey, 1982), preposition-noun pairs with respect to case (Lukatela, Kostić, Feldman, & Turvey, 1983), and possessive adjective-noun pairs with respect to gender (Gurjanov, Lukatela, Lukatela, Savić, & Turvey, 1985). To date, these grammatical congruency effects have been defined over the difference between congruent and incongruent situations and have not employed a neutral baseline. Relative amounts of facilitation and inhibition are not known.

These low constraint syntactic context effects are germane to the current discussion because they have been interpreted within a framework that is continuous with Forster's model of autonomous levels. The outputs of each level are considered to be available to the decision making device. In the normal course of language comprehension, all of these outputs are important and the processor heeds all of them. When the processor becomes specialized for lexical decision, it cannot obviate this characteristic. That is to say, even though lexical decision needs output from the lexical processor alone, the other subprocessors cannot be disengaged; their outputs—in the form of syntactic and pragmatic coherence checks—bias the decision making device. A positive bias, as when the context and target are grammatically congruent or pragmatically plausible, hastens lexical decisions relative to a negative bias, as when the context and target are an ungrammatical or implausible combination.

It is important to note that, in contrast to associative priming, these context effects are post-lexical. They do not change the speed with which a lexical entry is found. And they allow a form of *automatic* processing (deGroot, Thomassen, & Hudson, 1982) that is different from the spreading activation assumed to operate in the lexicon. If information needed for the coherence evaluation is provided in the lexical entries for context and target, then low constraint contexts (e.g., minimal grammatical contexts, unfocused sentence contexts) can have a facilitating (or, unlike spreading activation, an inhibiting) influence on lexical decision times without entailing the unlikely assumption that broad classes of items in the lexicon—for example, all feminine singular nouns—are activated or attended to.

One word contexts are useful because they allow tight control on the context-target associative relationship (e.g., it cannot accumulate insidiously from several words in the context) and on the stimulus onset asynchrony (SOA). This last benefit is of importance because in contrast to spreading activation, which decays over time, post-lexical coherence checks should be indifferent to the interval between context and target. Their output is simply "coherent" or "not coherent" and this will not change over time (although, presumably, there is an upper limit after which the context and target will no longer be considered as part of the same situation). Whatever pattern of facilitation and inhibition is found at a short SOA, therefore, should be repeated at a long SOA.

The situations to be explored in the present experiments are low constraint, single word semantic contexts. Grammatically congruent, semantically plausible adjective-noun pairs and grammatically congruent, semantically implausible adjective-noun pairs will be evaluated relative to xxx-noun baselines.[1] A positive bias from both the syntactic and message processors should produce facilitation relative to the neutral baseline. But a positive bias from the syntactic processor coupled to a negative bias from the message processor should effectively cancel each other, making that condition no different from a neutral context. Experiment 1 will investigate this contrast at an SOA of 300 ms and Experiment 2 will use an SOA of 800 ms.

---

[1] DeGroot et al. (1982) warn that the xxx baseline may, in fact, be inhibitory and that a more neutral context is provided by a word such as "blank." Because the Serbo-Croatian language is inflectional, however, all words are marked for a grammatical role. Consequently, almost any seemingly neutral word would necessarily facilitate those words with which it was grammatically congruent and inhibit those with which it was incongruent. An exception is provided by noun contexts for noun targets—such pairs do not create a syntactic situation (Lukatela & Popadić, 1979)—but these introduce the possibility of associative or semantic relatedness. While the concerns of deGroot et al. (1982) are important, it may be that in Serbo-Croat, xxx contexts are as neutral as it gets. It has been suggested that a high proportion of baseline trials may serve to limit the inhibitory influence of xxx contexts (deGroot et al., 1982). Both of our experiments follow this recommendation by including 50% baseline trials.

## Experiment 1

### Method

**Subjects.** Twenty-six students from the Department of Psychology in the Faculty of Philosophy at the University of Belgrade participated in the experiment in partial fulfillment of a course requirement. All subjects had previously participated in reaction time experiments.

**Materials.** Critical context-target pairs consisted of 26 congruous adjective-noun pairs (e.g., BELI GOLUB, "white pigeon") and 26 incongruous adjective-noun pairs (e.g., VUNENA ŠKOLA, "woolen school") drawn from the mid-frequency range (Dj. Kostić, 1965). Targets ranged from 4-7 letters in length. (Because associative norms do not exist for Serbo-Croat, possible associative relationships were eliminated on the basis of a pretest.) All pairs were in the nominative case. Half of the pairs (in both conditions) were feminine and half were masculine. In addition, 52 adjective-pseudonoun pairs were constructed in which the pseudonouns differed from real words by replacing one or two letters but preserving the inflectional ending so that the pairs would not be grammatically incongruent.[2] The adjectives were the same as those that had been paired with the nouns. Finally, 104 baseline pairs were constructed by appending a context of 3 crosses (xxx) to all of the nouns and pseudonouns.

**Design.** Each subject saw 26 adjective-noun pairs (half congruent and half incongruent), 26 adjective-pseudonoun pairs, 26 xxx-noun pairs, and 26 xxx-pseudonoun pairs. Subjects were randomly assigned to one of two counterbalancing groups as illustrated in Table 1. A given subject never encountered a given target or context (other than xxx) more than once.

**Table 1**

**Illustration of the Design and (Translated) Examples of Stimuli Used in the Experiments**

| | | **Context-target relation** | | | |
|---|---|---|---|---|---|
| Group | Noun Gender | Congruous | Incongruous | Neutral | Pseudoword |
| A | F | THIN-HAIR | SLEEPY-DOOR | XXX-AUNT | GOOD-GREEB |
| | M | DEEP-POT | SLOW-COAT | XXX-DEER | SPEEDY-CLUD |
| B | F | GOOD-AUNT | SOUR-CAT | XXX-HAIR | THIN-SPORL |
| | M | SPEEDY-DEER | HAPPY-NAIL | XXX-POT | DEEP-LORT |

**Procedure.** A subject was seated before the CRT of an Apple IIe computer in a dimly lit room. A fixation point was centered on the screen. On each trial, the subject heard a brief warning signal after which an adjective or xxx appeared for 300 ms centered above the fixation point. Immediately after the context disappeared (SOA of 300 ms) a noun or pseudonoun appeared below the fixation point for 1400 ms. All letter strings appeared in uppercase Roman. Subjects were instructed to decide as rapidly as possible whether or not the second stimulus was a word. To ensure that subjects were reading the contexts, they were occasionally asked to report both stimuli after the lexical decision had been made. Decisions were indicated by depressing a telegraph key with both thumbs for a "No" response or by depressing a slightly further key with both forefingers for a "Yes" response. Latencies were measured from the onset of the target. If the response latency was longer than 1500 ms, a message appeared on the screen requesting that the subject respond more quickly. The experimental sequence was preceded by a practice sequence of 20 different context-target pairs.

---

[2] For pseudonouns following nominative adjectives, the pairs cannot be decisively congruent, though, because of the way in which case is marked in nouns. An inflection such as -A indicates nominative for feminine singular nouns but genitive for singular masculine nouns. That is, access of the lexicon is required in order to render the inflection unambiguous.

### Results

Latencies in excess of 1500 ms and less than 350 ms were excluded from the analysis. The means of the subjects' latencies are shown in Figure 1 and their percentage errors (wrong and slow responses) are presented in Table 2 (None of the error analyses revealed any significant differences). A prime x congruence ANOVA on the acceptance latencies revealed a main effect of prime, $F(1,25) = 8.04, MSerr = 1909.44, p < .01$ (word primes averaged 674.5 ms while xxx primes averaged 699 ms) and congruence, $F(1,25) = 5.54, MSerr = 2452.07, p < .03$ (congruent situations averaged 675.5 ms, while incongruent xxx primes averaged 698 ms). The prime x congruence interaction was significant, $F(1,25) = 28.85, MSerr = 1083.95, p < .001$. Protected t-tests (Cohen & Cohen, 1975; the error term from the ANOVA is used as the estimate of the variance) were conducted on the means for congruous versus baseline, $t(25) = 4.87, p < .01$, and incongruous versus baseline, $t(25) = .82, p > .10$. In other words, there was facilitation but no inhibition.

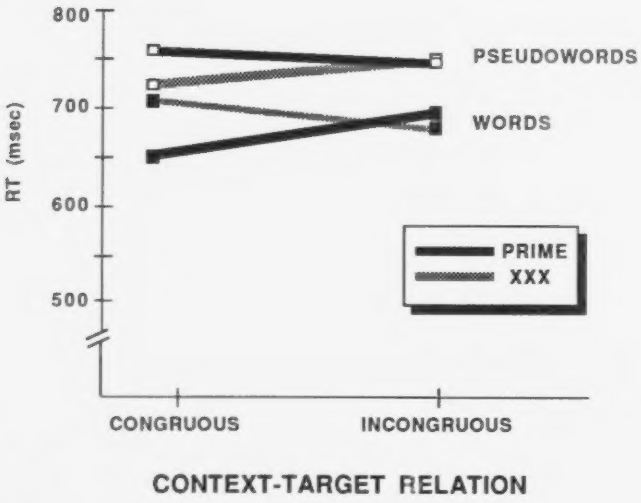

**Figure 1.** Average lexical decision latencies to word and pseudoword targets as a function of the semantic relationship between context and target at an SOA of 300 ms (Experiment 1).

Table 2

Percentage of Incorrect Lexical Decisions for Semantically Congruous and Incongruous Pairs with an SOA of 300 ms

| | Words | | Pseudowords | |
|---|---|---|---|---|
| Context-target relation[a] | Prime | XXX | Prime | XX |
| Congruous | 1.18 | 2.07 | 1.48 | 1.18 |
| Incongruous | 2.37 | 2.66 | 2.37 | 0.59 |

[a] Labels are defined for words and applied to pseudowords with corresponding contexts.

The pattern of results was largely corroborated by the stimulus analysis of acceptance latencies. The effect of prime was again significant, $F(1,50) = 6.24, MSerr = 2588.71, p < .02$, but the effect of congruence was not, $F(1,50) = 3.51, MSerr = 3804.01, p < .07$. The interaction, $F(1,50) = 11.78, MSerr = 2588.71, p < .001$, revealed the same pattern of facilitation as was found in the subjects analysis: protected t-tests indicated that there was facilitation for congruous situations, $t(50) = 5.93, p < .01$, but not inhibition for incongruous situations, $t(50) = .93, p > .10$.

For the rejection latencies, there was no effect of congruence, $F < 1$. The effect of prime was significant, $F(1,25) = 10.91, MSerr = 891.73, p < .01$ (word primes averaged 744.5 ms; xxx primes averaged 725.0 ms). Their interaction was significant, $F(1,25) = 11.17, MSerr = 1175.95, p < .01$. Protected t-tests revealed inhibition of the "congruent" pseudowords, $t(25) = 5.07, p < .01$, but no effect on "incongruent" pseudowords, $t(25) = .30, p > .10$.

This was duplicated in the stimulus analysis of rejection latencies. Prime was significant, $F(1,50) = 5.05, MSerr = 2003.52, p < .03$, but congruence was not, $F < 1$. The interaction was again significant, $F(1,50) = 6.52, MSerr = 2003.52, p < .02$. Protected t-tests revealed inhibition in the congruous situations, $t(50) = 4.8, p < .01$, but no difference for incongruous situations, $t(50) = .30, p > .10$.

## Experiment 2

### Method

**Subjects.** Twenty-six students from the Department of Psychology in the Faculty of Philosophy at the University of Belgrade participated in the experiment in partial fulfillment of a course requirement. All had experience in reaction time experiments but none had participated in Experiment 1.

**Materials and design.** The same as Experiment 1.

**Procedure.** The same as Experiment 1 with the exception that the SOA was 800 ms.

## Results

Latencies in excess of 1500 ms and less than 350 ms were excluded from the analysis. The means of subjects' latencies are shown in Figure 2 and their percentage errors (wrong and slow responses) are presented in Table 3. A prime x congruence ANOVA on the acceptance latencies revealed significant differences of prime, $F(1,25) = 33.78, MSerr = 1082.71, p < .001$ (with word primes averaging 626.5 ms and xxx primes averaging 664 ms), congruence, $F(1,25) = 4.93, MSerr = 2482.99, p < .04$ (with congruent situations averaging 634.5 ms and incongruent situations averaging 656 ms), and a prime x congruence interaction, $F(1,25) = 17.92, MSerr = 1241.14, p < .001$. Protected t-tests revealed significant facilitation for congruous nouns, $t(25) = 7.34, p < .01$, but not inhibition for incongruous nouns, $t(25) = .87, p > .10$. The error analysis revealed an effect of prime, $F(1,25) = 10.21, MSerr = 10.92, p < .004$.

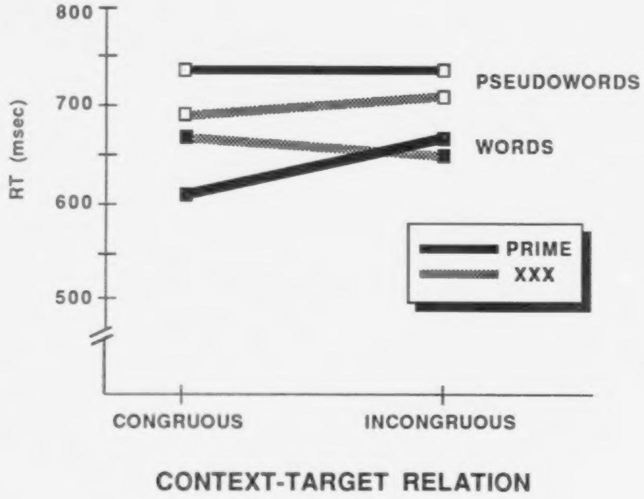

**Figure 2.** Average lexical decision latencies to word and pseudoword targets as a function of the semantic relationship between context and target at an SOA of 800 ms (Experiment 2).

For the rejection latencies, there was an effect of prime, $F(1,25) = 12.55, MSerr = 1572.83, p < .002$ (word primes averaged 728.5 ms, xxx primes averaged 701 ms) but neither congruence, $F(1,25) = 3.34, MSerr = 1286.6, p < .08$, nor the interaction, $F < 1$, reached significance. No differences were found by the error analysis.

In the stimulus analysis of acceptance latencies the effect of congruence was not significant, $F(1,50) = 2.49, MSerr = 4351.19, p > .10$. The main effect of prime, $F(1,50) = 12.99, MSerr = 2925.91, p < .001$, and the interaction, $F(1,50) = 8.29, MSerr = 2525.91, p < .01$, were significant. The error analysis showed an effect of prime, $F(1,50) = 7.38, MSerr = 15.11, p < .01$. For rejection latencies, prime was significant, $F(1,50) = 8.19, MSerr = 2589.09, p < .01$. Neither the effect of congruence nor the interaction reached significance, $F < 1$. No significant differences were found in the error analysis.

Table 3

**Percentage of Incorrect Lexical Decisions for Semantically Congruous and Incongruous Pairs with an SOA of 300 ms**

| | Words | | Pseudowords | |
|---|---|---|---|---|
| Context-target relation[a] | Prime | XXX | Prime | XX |
| Congruous | 0.89 | 2.07 | 2.66 | 0.89 |
| Incongruous | 1.18 | 4.14 | 2.07 | 1.48 |

[a] Labels are defined for words and applied to pseudowords with corresponding contexts.

## Discussion

As expected, plausible low constraint semantic contexts produced a facilitatory effect on word recognition while implausible low constraint semantic contexts yielded lexical decision times that were not different from a neutral baseline. The lack of inhibition for incongruous situations would not be predicted by a generalized priming story (e.g., Schwanenflugel & Shoben, 1985). This is particularly true at the longer SOA (cf. Neely, 1977) where the effect of attentional processes ought to be greater. Indeed, Becker (1980) has demonstrated inhibition dominance when the set of expected targets is not narrow. This latter result was obtained with associates (where the context was a category and target could be a typical or nontypical member of that category), however, and would not have tapped the semantic plausibility of a particular pair. We conjecture that the lack of inhibition in the semantically implausible situations in the present experiments derived from the fact that, because all situations were grammatically congruent, a positive bias from the syntactic coherence check cancelled the negative bias from the message level coherence check. The resulting situation was equivalent to having no context.[3]

Superficially, it might seem that the pseudoword data, which generally showed inhibition relative to the baseline, contradict this interpretation: Why isn't negative bias from the message processor cancelled by positive bias from the syntactic processor? We suspect that, because of the way in which case is marked in nouns, the syntactic processor is put into a "holding pattern," giving neither negative nor positive bias. Negative bias is absent because the syntactic relationship of the adjective-pseudonoun pairs is not immediately suspect. A negative bias would occur if the pseudonoun's inflection unambiguously indicated that its case was inappropriate for the preceding adjective (e.g., BELI BRAKU is unequivocally incongruent because the nominative adjective is followed by a pseudonoun marked for the accusative case). But such situations were not used here. Nonetheless a positive bias cannot be given either, because the inflections with which pseudonouns were constructed were ambiguous. For example, whether -A indicates that a singular noun is genitive (and, therefore, incongruent) or nominative (and, therefore, congruent) depends on the noun's gender (see Footnote 2). The problem arises because, for nouns, gender information is obtained from the lexicon, not the surface morphology of the letter string. There is, of course, no lexical entry for pseudonouns. This means that the syntactic processor continues to run, waiting for the information it needs to evaluate these syntactic situations. It would most likely be stopped only when a general system-level decision deadline is reached (cf. Coltheart, Davelaar, Johansson, & Besner, 1977). In contrast, xxx contexts should not engage the syntactic processor; the situations they create should be recognized as situations not requiring syntactic evaluation. When xxx-pseudoword (or xxx-word) is encountered, therefore, the syntactic processor makes no attempt to assign a syntactic structure to it. Decision time in nonsyntactic contexts can be influenced simply by the lexical processor, yielding faster responses than when the syntactic processor is caught up "waiting for Godot."

---

[3] Because association norms have not been compiled for the Serbo-Croatian language, one might argue that the experimental materials were, in fact, weak associates and nonassociates rather than low constraint plausible and implausible contexts. If this were the case, however, then we should expect no effect on the former and inhibition on the latter (cf. deGroot et al., 1982).

The question of whether or not the syntactic processor is engaged during the experimental situation also speaks to the difference between the present pseudoword data, where xxx contexts have a facilitating effect relative to word contexts, and other research, where the neutral context has an inhibiting effect (e.g., Balota, 1983; deGroot et al., 1982; Neely, 1976, 1977). As noted, the adjective-noun and adjective-pseudonoun situations used in the present investigation involved syntactic as well as semantic relations. More commonly, noun-noun associative pairs are employed and these appear not to be treated as syntactic situations (e.g., two semantically unrelated nouns that are in the same case do not show facilitation relative to those same nouns in incongruent cases [Lukatela & Popadić 1979]). The difference between unrelated word contexts and xxx contexts is, as deGroot et al. (1982) have argued, attributable to the inhibiting influence of xxx. In the present experiments, that inhibiting influence was either nullified by the high proportion of xxx trials (see Footnote 1) or counteracted by the futile attempts at a syntactic evaluation.

Further support for an interpretation in the framework of autonomous coherence checks comes from the duplication of the facilitation pattern at the short and long SOAs. The amount of facilitation was similar—59 ms at SOA 300 ms and 67 ms at SOA 800 ms—and the amount of inhibition was small and not significant at either interval. In contrast to a priming account, it can be argued that congruence effects defined at the syntactic or message levels ought to be rate-independent. Because the processing takes the form of a coherence evaluation with simply a positive or negative result, there is no avenue for time (other things being equal) to influence the outcome of the evaluation. The overall hastening of lexical decision from 300 ms SOA to 800 ms SOA (by 42 ms for words and 20 ms for pseudowords) is likely to be a general result of preparatory processes common to reaction time tasks (Gottsdanker, 1980) rather than an indication of a change in language processing at the two intervals.

It would be useful to investigate the time course of low constraint facilitation in a naming task as comparisons of lexical decision and naming are often informative (cf. West & Stanovich, 1982). In studies of associative priming, for example, deGroot (1984, 1985) has found that facilitation of lexical decision does not increase significantly over SOAs but facilitation of naming does. She suggests that "meaning integration" (the message processor) overshadows the effect of context-induced attentional processing in lexical decision but in naming, which does not engage the message level, the effect of attention can be seen to increase over SOAs. Failures to date to find semantic priming of naming in Serbo-Croat (Katz & Feldman, 1983), however, prohibit such a comparison here. Lupker (1984) has pointed out that so-called semantic priming actually hinges on the associative relationship between the context and target. If this is controlled for completely, then purely semantic relationships would produce no facilitation. Comparing strong and weak associates in a naming task would not address the issue of facilitation by low constraint *semantic* contexts.

Nonetheless, the present results are consistent with a number of experiments that exploit the inflectional nature of Serbo-Croat in investigations of syntactical processing. Neither spreading activation nor a prelexical attentional type of priming is supported by a pattern of findings that militate strongly for post-access coherence checks. We will summarize the argument here but see Gurjanov et al. (1985b) for the complete line of reasoning. As already mentioned, the standard result is that the target in a grammatically congruent pair is evaluated more quickly than the target in a grammatically incongruent pair (e.g., Gurjanov et al., 1985a, 1985b; Lukatela et al., 1982, 1983). Of particular interest is the fact that the magnitude of the grammatical congruency effect for adjective-noun pairs is matched by that found for pseudoadjective-noun pairs, both in visual (Gurjanov et al., 1985b) and auditory lexical decision (Katz, Boyce, Goldstein, & Lukatela, 1987). The observed influence of a pseudoadjective on the processing of a noun could only have been achieved through a relating of their respective inflections. The information required in order that a syntactic device might evaluate such relations is of three kinds: (1) inflections must be distinguished from stems; (2) word class must be identified; and (3) word gender must be identified. These three kinds of information are made available by lexical access.

What is the theoretical significance of low-constraint facilitation of word recognition? As the argument is usually developed, such effects are supposed to infirm models of autonomous processing because such effects imply that high level information is interacting with low level processes. In their summary of the issue, Sanocki et al. (1985, p. 147) observe:

> A facilitatory effect of low-constraint contextual information would be of particular theoretical interest, because it would implicate a linguistically powerful mechanism... A facilitatory effect of such a context would implicate a high-level mechanism that could affect more words than word level mechanisms (e.g., Becker, 1980; Neely, 1977) could affect.

Forster, architect of perhaps the strongest autonomous model, also sees low constraint sentences in the same light: "This theory clearly requires that sentence contexts should not influence lexical processing (either positively or negatively)" (1981, p. 471). We agree that a model of autonomous processing cannot accommodate such effects *on lexical processing*, but we do not agree that the existence of low constraint context effects necessarily implies the existence of "a linguistically powerful mechanism" that is, indeed, influencing lexical processing. Rather, the message processor does its evaluation on the basis of *information available in the lexical entries* of the accessed words. As Forster (1979) has pointed out, this may require a reconceptualization of the kind of information that is thought to be contained in the lexicon. The automaticity of sentence context effects—especially as evidenced by their stability over SOAs—may demand such a reconceptualization.

In the model advocated here, sentence context effects arise because of the integrity of the language processor, which cannot short circuit its own style of normal language comprehension. That is, the decision making device ordinarily must use the outputs of all three subprocessors in order to understand sentences. Negative bias from any level may be "a signal that perception or comprehension has failed and that some reanalysis is called for" (Fischler & Bloom, 1979,

p. 224; see also Kinoshita, Taft, & Taplin, 1985). For example, one might be alerted to an unfamiliar or inappropriate word or to a questionable syntactic construction (e.g., is a double negative intentional?). These effects are decidedly post-lexical but they are no less automatic because of it.

## References

Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning and Verbal Behavior, 22*, 88-104.

Becker, C. A. (1979). Semantic context and frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 252-259.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8*, 493-512.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI*. London: Academic Press.

deGroot, A. M. B. (1984). Primed lexical decision: Combined effects of the proportion of related prime target pairs and the stimulus-onset asynchrony of prime and target. *Quarterly Journal of Experimental Psychology, 36A*], 253-280.

deGroot, A. M. B. (1985). Word-context effects in word naming and lexical decision. *Quarterly Journal of Experimental Psychology, 37A*, 281-297.

deGroot, A. M. B., Thomassen, A. J. W. M., & Hudson, P. T. (1982). Associative facilitation of word recognition as measured from a neutral prime. *Memory & Cognition, 10*, 358-370.

Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior, 18*, 1-20.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. Walker (Eds.), *Sentence processing*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous processing. *Quarterly Journal of Experimental Psychology, 33A*, 465-495.

Gottsdanker, R. (1980). The ubiquitous role of preparation. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 355-371). Amsterdam: North-Holland.

Gurjanov, M., Lukatela, G., Lukatela, K., Savić, M., & Turvey, M. T. (1985a). Grammatical priming of inflected nouns by possessive adjectives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 692-701.

Gurjanov, M., Lukatela, G., Moskovljević, J., Savić, M., & Turvey, M. T. (1985b). Grammatical priming of inflected nouns by inflected adjectives. *Cognition, 19*, 55-71.

Katz, L., Boyce, S., Goldstein, L., & Lukatela, G. (1987). Grammatical information effects in auditory word recognition. *Cognition, 25*, 235-263.

Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 157-166.

Kinoshita, S., Taft, M., & Taplin, J. E. (1985). Nonword facilitation in a lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 346-362.

Kostić, Dj. (1965). *Frequency of occurrence of nouns in Serbo-Croatian*. Unpublished manuscript, Institute of Experimental Phonetics and Speech Pathology, University of Belgrade.

Lukatela, G., Kostić, A., Feldman, L. B., & Turvey, M. T. (1983). Grammatical priming of inflected nouns. *Memory & Cognition, 11*, 59-63.

Lukatela, G., Moraca, J., Stojnov, D., Savić, M., Katz, L., & Turvey, M. T. (1982). Grammatical priming effects between pronouns and inflected verb forms. *Psychological Research (Psychologische Forschung), 44*, 297-311.

Lukatela, G., & Popadić, D. (1979). Priming nouns by nouns. Unpublished manuscript. Laboratory of Experimental Psychology, University of Belgrade.

Lupker, S. J. (1984). Semantic priming without association: A second look. *Journal of Verbal Learning and Verbal Behavior, 23*, 709-733.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilatory and inhibitory processes. *Memory & Cognition, 4*, 648-654.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226-254.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55-85). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sanocki, T., Goldman, K., Waltz, J., Cook, C., Epstein, W., & Oden, G. C. (1985). Interaction of stimulus and contextual information during reading: Identifying words within sentences. *Memory & Cognition, 13*, 145-157.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 232-252.

West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 5*, 385-399.

Wright, B., & Garrett, M. (1984). Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition, 12*, 31-45.

# THE USE OF MORPHOLOGICAL KNOWLEDGE IN SPELLING DERIVED FORMS BY LEARNING-DISABLED AND NORMAL STUDENTS

**Joanne F. Carlisle**†

**Abstract.** *Currently popular systems for classification of spelling words or errors emphasize the learning of phoneme-grapheme correspondences and memorization of irregular words, but do not take into account the morphophonemic nature of the English language. This study is based on the premise that knowledge of the morphological rules of derivational morphology is acquired developmentally and is related to the spelling abilities of both normal and learning-disabled (LD) students. It addresses three issues: 1) how the learning of derivational morphology and the spelling of derived words by LD students compares to that of normal students; 2) whether LD students learn derived forms rulefully; and 3) the extent to which LD and normal students use knowledge of relationships between base and derived forms to spell derived words (e.g., "magic" and "magician"). The results showed that LD ninth graders' knowledge of derivational morphology fell between that of normal sixth and eighth graders, following similar patterns of mastery of orthographic and phonological rules, but that their spelling of derived forms was equivalent to that of fourth graders. Thus, they know more about derivational morphology than they use in spelling. In addition, they were significantly more apt to spell derived words as whole words, without regard for morphemic structure, than even the fourth graders. Nonetheless, most of the LD spelling errors were phonetically acceptable, suggesting that their misspellings can not be attributed primarily to poor knowledge of phoneme-grapheme correspondence.*

## Introduction

In order to gain insight into the nature of spelling abilities and disabilities, we must have an approach to classifying words and/or spelling errors that reflects a model of the spelling process

---

† Also American International College

*Acknowledgment.* I am indebted to Laurel Fais and students in the Language Training program at the Forman School in Litchfield, Connecticut, for their participation in this study. I would like to thank Isabelle Y. Liberman, Ignatius Mattingly, Donald Shankweiler, and Lee Sirois for their advice. The first phase of this research project was sponsored by NICHD grant HD-01994 to Haskins Laboratories and by a Dissertation Fellowship from the University of Connecticut.

and hypotheses about the nature of spelling disabilities. Currently, the most popular model of the process of spelling includes two distinct systems for spelling a word—a "whole word" system, which is dependent on recall of the word as a gestalt, and a "correspondence" system, which is dependent on knowledge of the ruleful relationships between sounds and letters. While this dual-system model, which can be termed a "phonetic"/"nonphonetic" model, has provided insight into certain aspects of spelling disabilities, it does not take into account the morphemic structure of words. For a complete understanding of the linguistic deficits of disabled spellers, we must take into consideration students' acquisition of morphological knowledge, as well as their ability to use this knowledge in spelling.

Described by a variety of terms (e.g., "regular" and "irregular," or "predictable" and "unpredictable"), the "phonetic"/"nonphonetic" approach has become the theoretical basis for extensive research on and diagnostic analysis of spelling disabilities (Barron, 1980; Boder, 1973; Boder & Jarrico, 1982; Camp & Dolcourt, 1977; Carpenter & Miller, 1982; Cook, 1981; Frith, 1980; Goyen and Martin, 1977; Holmes & Peper, 1977; Jorm, 1981; Moats, 1983; Nelson, 1980; Sweeney & Rourke, 1978; Whiting & Jarrico, 1980). Although the results of these investigations are not completely consistent (see Holmes & Peper, 1977), they have resulted in a consensus that learning-disabled or dyslexic spellers are apt to have a primary deficit that corresponds to one of the two systems—"phonetic" spelling or memory for "nonphonetic" words. Perhaps as a result, the "phonetic"/ "nonphonetic" distinction has been used as the basis for diagnostic tests that have become popular in the last ten years, including Larsen and Hammill's Test of Written Spelling (1976) and Boder's Test of Reading-Spelling Patterns (Boder & Jarrico, 1982). Boder (1973) argues that the prevalence of one of the two error types ("phonetic" and "nonphonetic") can be used to classify dyslexics into subgroups. By this system, spellers who cannot render words with phonetic accuracy are classified as "dysphonetic" and those who do not recall the configuration and characteristic visual features of words are classified as "dyseidetic", although it is possible to have both kinds of deficit and be placed in a "mixed" category.

This method of diagnosing types of disabled spellers has several important shortcomings. First, the strict dichotomy requires that all words (or misspellings of words) be classified as either "phonetic" or "nonphonetic." Because any word that is not completely regular phonetically must be considered "nonphonetic," the class of words considered "nonphonetic" becomes very large and heterogeneous. In the Test of Written Spelling (Larsen & Hammill, 1976), "myself" and "everyone" are included in the list of "Unpredictable" words, even though each is a compound of two very common morphemes, "my" and "self," "every" and "one." In fact, these two words pose quite a different challenge for young spellers than other "Unpredictable" words on the same list, such as "music" and "campaign."

Second, the phonetic approach misrepresents the nature of our writing system. "Phonetic" spelling places emphasis solely on the phoneme as the unit of language, and analysis of words or spelling errors focuses on the letter or letters that can be used to spell each phoneme accurately. While knowledge of sound-to-letter correspondences and memorization of "nonphonetic" words are necessary, these are not the only sources of knowledge spellers need to bring to the task. For accurate spelling children must also use knowledge of grammatical structure and knowledge of orthographic and morphological patterns and rules, even in the first few years of school (Chomsky, 1970; Hanna, Hodges, & Hanna, 1971; Marino, 1979; Schwartz & Doehring, 1977).

Of specific concern here is the fact that the "phonetic"/"nonphonetic" system ignores the large role that morphemic structure plays in the formation of English words. The nature of our language is such that phonemes and morphemes are intricately embedded, so that the English language is accurately described as "morphophonemic" (Chomsky & Halle, 1968). In fact, analysis of errors at the "letter level" must be sensitive to students' knowledge of the structure of words to be meaningful. For example, in an analysis of errors made on "ie" words by junior high students (Carlisle & Liberman, 1983), transpositions of "ie" were found to be very common in words like "chief" and "belief," but nonexistent in words like "babies" or "parties." The reason may be that the linguistic role of "ie" in these words is quite different. The "ie" in "chief" falls within a single base morpheme, whereas the "ie" in "babies" occurs at the morphemic boundary, the point at which the plural marker "s" is added to the base "baby." Even the poorest spellers did not spell "babies" "babeis"; their misspellings of the "ie" were commonly "babes" and "babys". Ordinarily, analysis of "letter level" errors does not take into consideration students' knowledge of the morphemic structure of words.

While researchers believe that students must use morphological knowledge to be successful in reading and spelling (Chomsky, 1970; Hodges & Rudorf, 1966; Liberman, 1982; Venezky, 1970; Venezky & Weir, 1966), we know little about how children learn to use morphological knowledge, particularly in spelling. We know more about how inflected forms are learned than how derived forms are learned. By the age of seven, children generally use inflected forms rulefully in speaking (Berko, 1958; Selby, 1972). These forms include the verb tense markers (e.g., "-ed," "-ing"), the "s" plural and possessive markers, and so on. The derived forms are learned later and more slowly, starting with the more common regular forms such as "foggy" (the adjectival form of "fog") and "slowly" (the adverbial form of "slow") and progressing to forms that undergo phonological changes (as in "magic" and "magician") (Berko, 1958; Derwing, 1976; Derwing & Baker, 1979).

Learning derived forms is more difficult than learning inflected forms for several reasons. One reason is that inflected forms are more common, perhaps because they are necessary for the grammar of the language. Learning inflected forms is a more integral part of language acquisition than learning derived forms. In addition, while the phonological shifts from base to derived forms are often ruleful (Chomsky & Halle, 1968), they are complex and sometimes seemingly arbitrary. For example, "deep" becomes "depth," but "steep" does not become "stepth." Furthermore, word-specific knowledge seems to play a larger role in learning derived forms than in learning inflected forms (Klima, 1972; Smith & Sterling, 1982). Such knowledge includes the particular suffix used to form a given derived word. For example, formation of a noun from an adjective may be accomplished by adding on "-ness," "-ment," or "-ity." Sometimes two grammatically identical forms exist in the language, varying only slightly in meaning (e.g., "bountiful" and "bounteous"). Linguistic rules do not consistently specify the exact forms of derived words found in the language.

Learning to read is believed to help the child acquire the derived forms as patterns or word families. The orthography preserves the identity of the word, even when phonological changes take place (e.g., "equal," "equality"). In addition, some orthographic shifts can be learned as patterns (e.g., "divide" and "division," "decide" and "decision") (Chomsky, 1970; Templeton, 1980). It is not surprising, then, that good readers have been shown to have a more thorough knowledge of derived forms than poor readers (Barganz, 1971; Freyd & Baron, 1982).

Children's ability to spell derived forms has received less attention. We know that children begin to learn the patterns of morphemically complex words in their first years in school (Schwartz & Doehring, 1977). For instance, as early as first grade, linguistically mature students spell words that sound alike (e.g., "wind" and "pinned") in ways that reflect differences in morphological structure (Rubin, 1984). Still, while these early studies suggest that spelling of inflected forms is learned rulefully, they do not speak directly to the issue of how children go about spelling derived forms. It is possible that derived forms are spelled as whole words, without reference to their morphological structure. Support for this position comes from Sterling (1983), who has found patterns of errors indicating that inflected forms are learned rulefully, while derived forms are learned as whole and independent words. The alternative is that some spellers, at least, spell derived words by making use of knowledge of the morphemic structure of the word. We might suspect that better spellers would make superior use of knowledge of derivational morphology than poor spellers. There is some evidence to support this hypothesis. Several researchers (Fischer, Shankweiler, & Liberman, 1985; Templeton, 1980; Templeton & Scarborough-Franks, 1985) have provided evidence that good spellers, particularly at high school and college levels, have superior knowledge of phonological and orthographic rules.

Poor spellers may lack linguistic knowledge, but their weaknesses are not just at the level of representing phonemes. We have evidence that poor spellers spell inflected and derived words with a high degree of phonetic accuracy but have difficulty adding suffixes to base words accurately (Carlisle, 1984). We do not know whether they lack morphological knowledge or simply the ability to use that knowledge in spelling. In a study of the spelling of good and poor junior-high spellers, some students wrote "easally" for the word "easily," given the sentence, "Our team won the race ____ ." And some wrote "finely" for "finally," given the sentence, "I have ____ finished my lesson." We do not know whether these students know that "final" is the base word of "finally" or that "ease" and "easy" are in the same word family. In fact, to understand such spelling errors, we must know whether students at this level lack knowledge of morphological relationships, or whether they do not think to use this knowledge in spelling derived words.

The design of the present study reflects the belief that in order to understand the full range of spelling capabilites of disabled spellers, we need to learn more about the knowledge of the morphemic structure of both normal and disabled spellers. In an earlier study, students in the fourth, sixth, and eighth grades were selected to investigate the normal developmental learning of derivational morphology and the ability to spell derived forms. For the present study, a group of learning-disabled ninth-grade students with spelling disabilities were selected for comparison to the normal students. The ninth-grade level was chosen in light of the findings of previous studies indicating that dyslexic or learning-disabled students were commonly three to five years delayed in their acquisition of spelling skill and morphological knowledge (Moats, 1983; Wiig, Semel, & Crouse, 1973). Thus, it was estimated that the ninth-grade LD students might developmentally resemble the fourth or sixth graders in the acquisition of derivational morphology and the spelling of derived words.

Initially, a study was undertaken to investigate 1) the developmental learning of derivational morphology and its rule systems (phonological and orthographic rules) by normal children in grades four, six, and eight and 2) the extent to which these students use knowledge of morphological relationships in their spelling of derived words. The purpose of the present study was to determine the extent to which LD students' learning of derivational morphology and spelling of derived words differed from that of the normal students. This study was designed to address three questions: First, do LD students know and use rules of derivational morphology in the same way as do peers at a similar level of spelling ability? Second, do the LD students appear to be learning the underlying phonological and orthographic rules of derivational morphology? And, third, do LD students use their knowledge of the morphemic structure when they spell derived words?

## Method

The description of the present study has included the normal groups (fourth, sixth, and eighth graders) of the first study (Carlisle, 1985) for purposes of comparison. The study was designed to determine whether learning-disabled (LD) students showed similar or different patterns of learning derivational morphology and spelling derived forms.

## Subjects

The normal students were fourth, sixth, and eighth graders who were members of classes studying reading or language arts in a rural school system. There were 22 fourth graders, 22 sixth graders, and 21 eighth graders; all students were reported by their teachers to have normal intelligence. The LD students were ninth graders who attended a rural private high school with a specific program of remedial training for LD students. The 17 students who participated were all previously evaluated and determined to have specific learning disabilities in reading and written language skills. The mean intelligence quotient of these students was reported by the school to be 107.

The Wide Range Achievement Spelling subtest (Jastak & Jastak, 1978) was used to compare the groups on spelling ability. As Table 1 shows, the LD ninth graders' mean score closely resembled that of the fourth graders. The LD ninth graders' performance did not differ significantly from that of the fourth graders, $t(37) = 0.08, p = 0.937$, but did differ significantly from that of the sixth graders, $t(37) = 2.14, p < .05$, and the eighth graders, $t(36) = 8.99, p < .001$.

**Table 1**

**Performance on Wide Range Achievement Test (WRAT) Spelling by Grade Level**

| | Mean GE (and SD) | Subtest, Range |
|---|---|---|
| 4N | 5.9<br>(1.0) | 3.9 - 8.1 |
| 6N | 6.7<br>(1.4) | 3.9 - 8.9 |
| 8N | 9.4<br>(1.3) | 6.7 - 10.9 |
| 9LD | 5.9<br>(1.2) | 3.6 - 8.1 |

## Instruments

The following tests were administered:

1) The Wide Range Achievement Test (WRAT), Spelling subtest (Jastak & Jastak, 1978): This standardized spelling test was used to determine the spelling abilities of the four groups and to determine the validity of the experimental Spelling Test. The correlation between performance on the WRAT Spelling Test and the Spelling Test, Derived Forms subtest, was $.74(p < .001)$ for the fourth, sixth, and eighth graders.

2) The Test of Morphological Structure (TMS): This is a test of oral generation designed to assess knowledge of derivational morphology. It has two subtests, each with 40 items. The Derived Forms subtest requires that the student provide the appropriate derived form, given the base form of the word and a short sentence. The Base Forms subtest required the student to supply the base form, given the derived form and a short sentence. In both cases, the word the student supplied was the final word of the sentence. For example, the first item on the Derived Forms subtest is: "Warm. He chose the jacket for its —." The target response is "warmth." The first item on the Base Forms subtest is: "Growth. She wanted the plant to —." The target response is "grow."

The words on this test reflect four types of relationship in the transformation from base to derived forms. These are as follows: No Change in phonology or orthography (for example, "enjoy to enjoyment"); Orthographic Change only (for example, "sun" to "sunny" or "rely" to "reliable"); Phonological Change only (for example, "magic" to "magician" or "sign" to "signal"); and, Both Changes, orthographic and phonological (as in "deep" to "depth" or "decide" to "decision") (see Carlisle, 1985, for further description of the construction of this test). The ten base words included under each type of transformation were equated for word length and word frequency on both subtests of the TMS (Base Forms and Derived Forms) (Carroll, Davies, & Richman, 1971). The same procedure was used to equate the derived words under each type of transformation on each TMS subtest for word length and word frequency. The test was administered by a tape-recording of a native American male speaker.

3) The Spelling Test (ST): This experimental test is a test of dictated spelling, consisting of two parts—a Derived Forms and a Base Forms subtest, each with forty items. The student was presented with the word, a sentence containing the word, and then the word again. For example, the first item of the Derived Forms subtest is: "Explanation. The explanation was long. Explanation."

The words on the ST are the same words (base and derived forms) that comprise the Derived Forms subtest of the TMS; altogether there are forty pairs of words. Including pairs of base and derived forms allows for analysis of students' use of morphological knowledge in spelling. If a derived word is spelled by reference to its morphemic structure, a prerequisite must be the ability to spell the base form correctly. Alternatively, if the spelling of each of the two forms (base and derived) is learned independently (i.e., as whole words), we would expect that in some cases the derived form would be spelled correctly while the base form would be misspelled. Thus, the ST was constructed to examine the extent to which successful spelling of a base form was related to successful spelling of its derived counterpart. The test was administered by a tape-recording of a native American male speaker.

4) The Test of Suffix Addition (TSA): This experimental test is a paper-pencil task that required the students to combine a base word and a suffix, following the rules that govern the addition of suffixes to words. The test was designed to explore students' knowledge of the orthographic transformations between base and derived words. There are 30 items on the test. The base words are nonsense words, made by changing one consonant or consonant blend of a real word. The suffixes are real. For example, the first item is as follows: 1. dun + y = ______." Nonsense words were used in order to have a relatively pure test of the students' ability to apply suffix addition rules. The students could not simply know how to spell the whole word. Knowledge of three orthographic rules was evaluated— those governing the addition of suffixes to words ending in silent "e," to words ending in "y," and to words ending in a single consonant.

### Procedures

In both phases of the study, the students were administered the Wide Range Achievement Test (WRAT), Spelling subtest, and the three experimental tests described above—1) the Test of Morphological Structure (TMS), 2) the Spelling Test (ST), and 3) the Test of Suffix Addition (TSA). First, the WRAT, Spelling subtest, and the ST, Derived Forms subtest, were administered to each grade-level group. Between two to three weeks later, the ST, Base Forms subtest, and the TSA were administered to each grade-level group. (The Derived Forms subtest of the ST was administered before the Base Forms subtest so that the students would not be given the advantage of practice in spelling the base forms prior to spelling the derived forms.) Between one and two weeks later, the TMS was administered to each student individually.

## Results

### Performances of LD and Normal Students on the Experimental Tests

The first research question asked how the learning of derivational morphology and spelling of derived words by LD ninth-graders compared with that of normal students. This question was addressed by examining the students' performances on the Test of Morphological Structure (TMS), the Spelling Test (ST) and the Test of Suffix Addition (TSA), as shown in Table 2. On the TMS, the normal students showed clear developmental trends in their generation of the base and derived words, while the LD ninth graders' performance fell between the sixth- and eighth-grade levels. An analysis of variance showed significant differences between the groups on both the Derived Forms subtest, $F(3,78) = 18.914, p < .001$, and the Base Forms subtest, $F(3,78) = 16.879, p < .001$. On the Base Forms subtest post hoc analysis (Scheffé, $p < .05$) revealed that significant differences existed between all of the groups (the direction of the difference is indicated by the symbol $<$) : $4N < 6N < 9LD < 8N$. On the Derived Forms subtest the LD students' performance did not differ significantly from that of the sixth graders: $4N < 6N = 9LD < 8N$ (Scheffé, $p < .05$).

Developmental trends in the ability to spell base and derived forms were evident from the normal students' performance on the two subtests of the ST, while the performance of the LD ninth graders resembled that of the fourth graders (see Table 2). An analysis of variance showed significant differences in performance of the groups on the Base Forms subtest, $F(3,78) = 20.424, p < .001$, and on the Derived Forms subtest, $F(3,78) = 27.963, p < .001$. A comparison of the performance of the groups (Scheffé, $p < .05$) indicated that on both the Base Forms subtest and the Derived Forms subtest, the LD ninth graders' performance did not

Table 2

Performance on Experimental Tests of Morphological Structure(TMS), Spelling (ST), and Suffix Addition (TSA): Means and SDs

| | TMS* | | ST* | | TSA** |
|---|---|---|---|---|---|
| | Derived | Base | Derived | Base | |
| 4N | 27.0 | 30.8 | 14.5 | 24.9 | 16.0 |
| | (5.6) | (6.9) | (9.7) | (9.3) | (4.0) |
| 6N | 32.2 | 35.6 | 26.0 | 34.2 | 17.9 |
| | (3.5) | (3.7) | (7.5) | (4.1) | (3.3) |
| 8N | 36.0 | 39.4 | 34.4 | 38.2 | 21.0 |
| | (2.1) | (0.7) | (5.3) | (3.0) | (3.7) |
| 9LD | 33.0 | 37.8 | 16.8 | 28.1 | 17.5 |
| | (3.2) | (2.1) | (7.1) | (5.8) | (4.9) |

*Maximum possible = 40
**Maximum possible = 30

differ significantly from that of the fourth graders: $4N = 9LD < 6N < 8N$. Performance on the TSA indicated a somewhat different developmental trend. Although an analysis of variance showed significant difference between the groups, $F(3,78) = 6.017, p < .001$, the fourth graders' performance did not differ significantly from that of the sixth graders, and the LD ninth graders did not differ significantly from that of either the fourth or sixth graders (Scheffé, $p < .05$). Thus, knowledge of the rules that govern the addition of suffixes improved significantly only between the sixth and eighth grades: $4N = 9LD = 6N < 8N$.

### Discriminating the Groups by the TMS and ST Subtests

While the above analyses indicated the group differences on the Derived Forms and Base Forms subtests of the TMS and ST, they left open the question of which subtests best differentiate the groups. To address this question, the students' scores on these four subtests were subjected to a stepwise discriminant function analysis. Table 3 shows the standardized canonical coefficients for the two significant functions that were generated. For the first function, the coefficients were high for the subtests that measure morphological knowledge (the TMS Base Forms and Derived Forms and the ST Derived Forms); this function accounted for 71.52% of the variance ($p < .001$). The second function, explaining an additional 24.21% of the variance, for a total of 95.73%, was barely significant ($p = 0.05$). The highest coefficient was on the TMS, Base Forms subtest. The first function reflects group differences in knowledge of derived morphology. The second function may reflect word knowledge or vocabulary development.

**Table 3**

**The Standardized Canonical Coefficiencts of the Stepwise Discriminant Function Analysis of the Subtests of the Test of Morphological Structure (TMS) and the Spelling Test (ST)**

| Subtests* | Function 1 | Function 2 |
|---|---|---|
| ST, Derived | 0.95735 | -0.62458 |
| TMS, Base | -0.57937 | 1.16453 |
| TMS, Derived | 0.70170 | 0.09103 |
| ST, Base | 0.01884 | -0.14143 |

*Subtests are given in order of entry in the analysis.

## Ruleful Learning of Derivational Morphology

The second question addressed by the study was whether LD students' learning of derivational morphology reflects the ruleful nature of the morphological transformations between base and derived forms. To investigate this issue, performance of the groups was analyzed on the basis of the four types of transformation from base to derived forms. The four types of transformations between base and derived forms—"No Change" (NC), "Orthographic Change Only" (OC), "Phonological Change Only" (PC), and "Both Orthographic and Phonological Changes" (BC)—were equally represented on the TMS subtests.

An analysis of variance showed that the four groups differed significantly in their performance on each of the transformations on the TMS Derived and Base subtests; the univariate F ratios were all highly significant (see Table 4). Of particular interest is the fact that the pattern of performance across word types was very similar for the four groups, as can be seen in Figure 1. These graphs illustrate several results of note. First, the students consistently made the most errors on words that undergo phonological change or both phonological and orthographic changes. Second, the LD ninth graders' pattern of performance on the different transformations was quite similar to that of the normal students, indicating a lag in their mastery of the transformations, but not a noticeably different pattern in their learning of the four types of transformations in derivational morphology.

## The Spelling of Base-Derived Word Pairs

The third question addressed by this study was whether LD students spell derived words with reference to their morphemic structure. Toward this end, the spelling of the base and derived word pairs on the ST were scored according to the four possible patterns of performance: Both Incorrect (e.g., "equl" and "eqalty"), Base Correct/Derived Incorrect (e.g., "equal" and "eqalty"), Derived

## Table 4

**Univariate F Ratios of the Transformations on the Base Forms and Derived Forms of the Test of Morphological Structure (TMS)**

| | F-Ratio** |
|---|---|
| **Base Forms** | |
| No Change | 7.788* |
| Orthographic Change | 6.559* |
| Phonological Change | 15.300* |
| Both Change | 11.850* |
| **Derived Forms** | |
| No Change | 9.719* |
| Orthographic Change | 9.224* |
| Phonological Change | 9.593* |
| Both Change | 19.560* |

*$p < .0005$
**With 3 and 78 degrees of freedom.

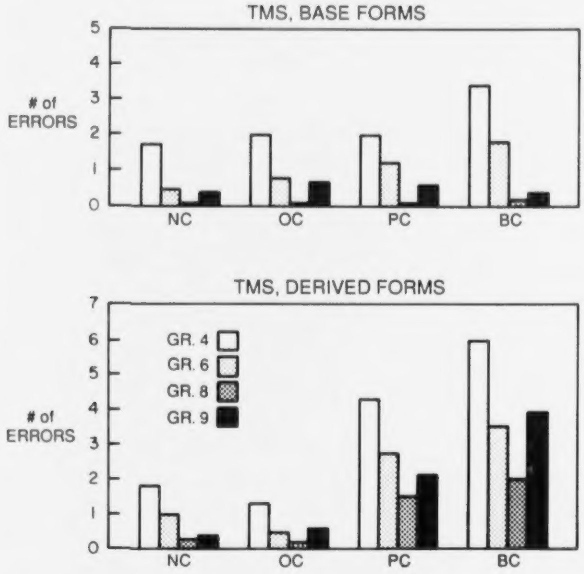

**Figure 1.** Mean errors on four types of transformation—No Change (NC), Orthograph Change (OC), Phonological Change (PC), and Both Change (BC)—on the Test of Morphological Structure (TMS) Base Forms and Derived Forms subtests.

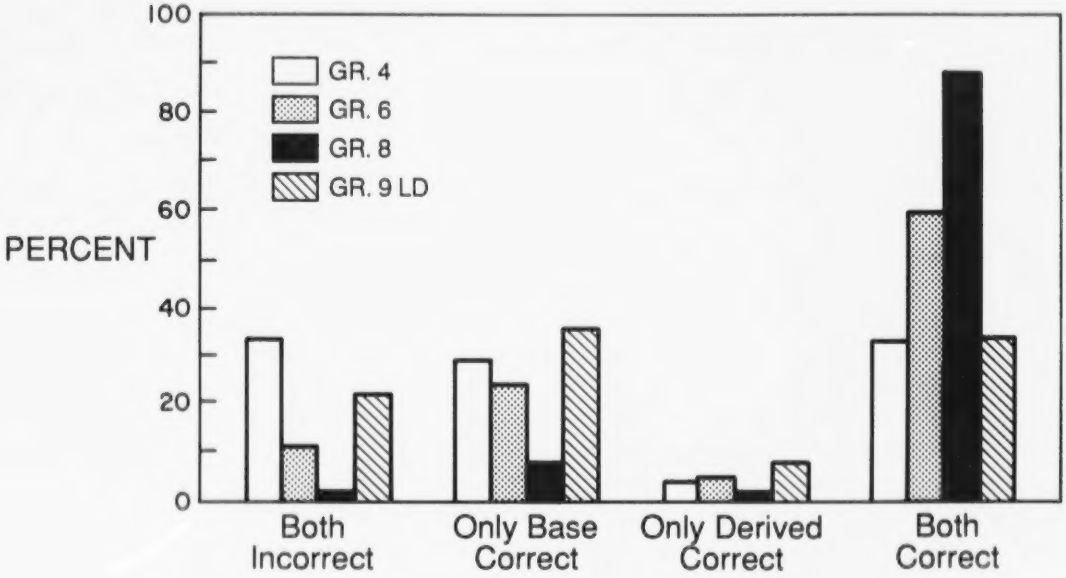

**Figure 2.** Spelling performance on pairs of base and derived words (expressed as % of opportunity).

Correct/Base Incorrect (e.g., "equality" and "equl"), and Both Correct ("equal" and "equality"). The proportion of overall performance for each pattern is given for each of the four groups in Figure 2.

Of particular interest are two of the categories—Base Correct/Derived Incorrect and Derived Correct/Base Incorrect, as they suggest the extent to which knowledge of the spelling of the base form is related to knowledge of the spelling of the derived form. An analysis of variance showed that the groups differed significantly on these two measures (Base Correct/Derived Incorrect, $F(3,78) = 24.414, p < .001$; Derived Correct/Base Incorrect, $F(3,78) = 11.303, p < .001$). Paired comparisons (Scheffé $p < .05$) indicated that the LD ninth graders had significantly more pairs that fell in the Base Correct/Derived Incorrect category than any of the other groups: $9LD > 4N > 6N > 8N$. Similarly, the LD ninth graders also had significantly more pairs that belonged to the Derived Correct/Base Incorrect pattern: $9LD > 4N = 6N > 8N$. Together, these findings indicate that the LD ninth graders more frequently spelled correctly ONE of the pair (base or derived word) than do the normal students, including the fourth graders.

## Discussion

Comparison of LD and normal students' performances on the tests of morphological knowledge and spelling of base and derived forms has confirmed several of the initial expectations. First, youngsters normally learn a great deal about derivational morphology between the fourth and eighth grades. The performance of the ninth-grade LD students suggests that while they are experiencing a lag in their mastery of derivational morphology, their pattern of learning the underlying phonological and orthographic rules resembles that of the normal students. Second, while both normal and LD students know more about morphological relationships that they use in spelling derived forms, the gap is more pronounced for the LD students. The normal students' spelling of base and derived word pairs shows that they spell many derived words by using knowledge of morphemic structure. This is not the case for the LD students. However, a post hoc examination of the LD students' spelling errors suggests that their difficulties spelling derived forms cannot be attributed solely to lack of mastery of phoneme-grapheme correspondence rules.

### The Learning of Derivational Morphology by Normal and LD Students

Understanding the patterns of performance by the normal students provided a reference by which to evaluate the performance of the LD students. Clear developmental trends were evident in both the oral generation of derived forms and the spelling of base and derived forms. Several points of particular interest might be emphasized here. First, on the Test of Morphological Structure (TMS), the students in all four groups consistently had an easier time when they were given the derived form (e.g., "growth") and were asked to supply the base form (e.g., "grow") than when they were given the base form (e.g., "warm") and were asked to supply the appropriate derived form (e.g., "warmth"). Extracting the base form is easier than generating the derived form. One of the central differences between the two tasks is that generating the derived form required some word-specific knowledge. Derivational rules cannot supply this particular kind of knowledge. Specific word knowledge helps us know that "equality," not "equalness," is the noun form of "equal." It is not surprising that the students' ability to generate the correct derived form lagged behind their ability to extract the base word. In fact, this pattern confirms our impression at the outset of this study that word-specific knowledge plays a large role in the level of learning of derivational morphology. It also shows, however, that rules governing the relationships between base and derived forms are learned. A second trend of interest is that spelling base and derived forms consistently lagged behind the ability to generate the same words. Spelling is evidently the more difficult task. As we discussed earlier, spelling draws on knowledge of sound-letter correspondences, syntactic roles, and orthographic rules as well as on knowledge of the morphology.

The particular concern of the present study is how the LD ninth graders compare to their normal peers in mastering derivational morphology and spelling derived forms. First, the LD ninth graders fell between the sixth and eighth graders on the TMS, resembling most closely the eighth graders in knowledge of base forms and the sixth graders in knowledge of the derived forms. In contrast, on the Base and Derived Spelling Test (ST) subtests, the LD ninth graders performed very much like the fourth graders. Thus, while they evidently are delayed in their acquisition of morphological knowledge, they are more seriously delayed in their mastery of the spelling of both base and derived words.

### Ruleful Learning of Derivational Morphology

Assessing the nature of the students' morphological knowledge was carried out to determine the extent to which learning about derivational morphology is ruleful. This analysis was an investigation of the number of errors on each type of transformation between base and derived forms—"No Change," "Orthographic Change," "Phonological Change," and "Both Changes." Performances on both subtests of the TMS showed that for all of the groups, the number of errors increased on the more complex transformations—that is, more errors were made on those word pairs that undergo phonological or both phonological and orthographic changes than on words that undergo no change at all or only an orthographic change. The error pattern across transformations is consistent on each grade level; there is no interaction between type of transformation and grade level. If ruleful learning did not take place, we would expect more or less equal numbers of errors on the four types of transformations by group and by subtest. Thus, the marked consistency of the pattern is a strong indication that the learning derivational morphology reflects the relative difficulty of learning the orthographic and phonological rules. The younger students know many more "No Change" pairs than "Phonological Change" pairs. Where both phonological and orthographic transformations occur between base and derived forms, learning of the relationship between base and derived forms is not complete even by the eighth grade.

## Spelling Base and Derived Word Pairs

The performance of the LD ninth graders resembled that of the fourth graders on the spelling of both the base and derived words. Examination of the spelling of the pairs of base and derived words on the ST showed that the normal students used knowledge of word structure in spelling the derived forms, but that the LD students were less apt to use such knowledge in their spelling of derived forms. When the pairs of words (each base and its derived forms) were examined for error patterns (see Figure 2), one pattern emerged for normal students at all three grade levels. The two components of this pattern were that 1) the higher the grade level, the fewer errors on both members of the pair, base and derived, and 2) the derived form was seldom spelled correctly if the base word was misspelled; or, put another way, the students rarely spelled only the derived word correctly. Clearly, for normal students, knowing how to spell the base form (e.g., "equal") probably precedes and aids in learning to spell the derived form (e.g., "equality"). For these students, then, knowledge of the morphemic components does appear to be used in spelling dictated words.

In contrast, the LD ninth graders were more apt to spell only one of the pair correctly, be it the base form or the derived form. This means that in some cases, the base word was spelled incorrectly (e.g., "glorry"), but the derived word was spelled correctly (e.g., "glorious"). The fact that the number of base incorrect/derived correct errors is significantly greater for ninth-grade LD students than for normal fourth graders suggests that they were more apt to spell derived forms as whole words, without regard for the relationship to the base form or the morphemic transformation. Thus, even though the LD ninth graders' overall performance on the ST was at the same level as the fourth graders', they nonetheless showed less evidence of using morphological knowledge in spelling derived forms.

It seemed important to consider the possibility that the LD students' spelling errors could be categorized in terms of the "phonetic"/"nonphonetic" dichotomy that is currently the most popular system for specifying spelling disabilities. A post-hoc tabulation of every spelling of every derived word on the ST, Derived Froms subtest, was carried out at each grade level. The misspellings were then analyzed by two judges to determine whether the misspellings were reasonable phonetic versions of the dictated word. The general finding was that only a small proportion of errors could be labeled phonetically unacceptable. As an example, Table 5 shows one of the "Phonological Change" words, "magician." By examining all of the versions of spelling this word, we see that almost all of the errors reflect difficulties learning the correct spelling of the suffix. As we noted earlier, the LD students were roughly equivalent to sixth graders in their knowledge of morphemic structure, but the misspellings illustrate that they were less able to use this knowledge in spelling. All but about four of the LD students' misspellings must be considered phonetically

Table 5

All Spellings of the "Phonological Change" Word, "magician"

| Grade: 4N<br>(n=22) | | 6N<br>(n=22) | | 8N<br>(n=21) | | 9LD<br>(n=17) | |
|---|---|---|---|---|---|---|---|
| magition | 5 | magician | 17 | magician | 16 | magition | 3 |
| magician | 3 | magican | 2 | magican | 2 | magician | 2 |
| magican | 3 | magision | 2 | magision | 2 | magicion | 2 |
| migion | 1 | magition | 1 | magition | 1 | magishion | 1 |
| migishon | 1 | | | | | migition | 1 |
| mjshier | 1 | | | | | magication | 1 |
| mudishon | 1 | | | | | midican | 1 |
| magish | 1 | | | | | magishan | 1 |
| magiton | 1 | | | | | meniton | 1 |
| magishion | 1 | | | | | migertion | 1 |
| smajison | 1 | | | | | majion | 1 |
| machishan | 1 | | | | | machishon | 1 |
| micgen | 1 | | | | | m—* | 1 |
| macian | 1 | | | | | | |

acceptable versions of the word. Thus, it seems that this group of LD students has acquired basic knowledge of sound-letter correspondences. Still, as the sixth and eighth graders' spelling of "magician" indicates, older and more capable spellers did not opt for the basic phonetic spellings (e.g., "shun" for "cian" in "magician"). They seem to have learned to override the process of direct phonetic representation when they have acquired productive understanding of morphemic structure of the words they spell. In contrast, when phonological transformations occur, the LD students were more apt to spell words phonetically, often without awareness of the relationship to the spelling of the base word.

In summary, this investigation of the spelling of derived words has found a noteworthy discrepancy between the LD students' ability to generate orally derived forms and their ability to spell derived forms. Spelling is clearly the more difficult task of the two, not only for the LD students but for the normal students as well. At all levels the students appear to know more about the morphemic structure of words than they use in their spelling. However, the gap between knowing derived words in spoken language and spelling them correctly is more pronounced for the LD students than it is for normal fourth, sixth, and eighth graders. This gap cannot solely be attributed to lack of understanding of basic phoneme-grapheme correspondences. Their misspellings, as a rule, are viable phonetic representations. Instead, they appear to lack awareness of the presence of base forms within derived counterparts, and they lack specific knowledge about how to spell suffixes and how to attach suffixes to base words correctly.

## References

Barganz, R. A. (1971). *The morphophonemic performance of good and poor readers* (Technical Report #182). Madison, WI: Research and Development Center for Cognitive Learning.

Barron, R. W. (1980). Visual and phonological strategies in reading and spelling. In U. Frith (Ed.), *Cognitive processes in spelling*. New York: Academic Press.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*, 150-177.

Boder, E. (1973). Developmental dyslexia: Prevailing diagnostic concepts and a new diagnostic approach. In H. R. Myklebust (Ed.), *Progress in learning disabilities* (Vol II). New York: Grune and Stratton.

Boder, E., & Jarrico, S. (1982). *The Boder Test of Reading-Spelling Patterns: A Diagnostic Test for Subtypes of Reading Disability*. NY: Grune and Stratton.

Camp, B. W., & Dolcourt, J. L. (1977). Reading and spelling in good and poor readers. *Journal of Learning Disabilities, 10*, 46-53.

Carlisle, J. F. (1984, March). *Mastery of the conventions of suffixation by good and poor spellers in grades seven through nine*. Paper presented at the New York Orton Dyslexia Society Conference, New York, NY.

Carlisle, J. F. (1985). The relationship between morphological knowledge and spelling ability in fourth, sixth and eighth graders. *Haskins Laboratories Status Report on Speech Research, SR-82-83*, 151-174.

Carlisle, J. F., & Liberman, I. Y. (1983, March). *The spelling of good and poor readers at the junior high level*. Paper presented at the Tenth Anniversary Conference of the New York Orton Dyslexia Society, New York, NY.

Carpenter, D., & Miller, L. J. (1982). Spelling ability of reading disabled LD students and able readers. *Learning Disabilities Quarterly, 5*, 65-70.

Carroll, J. B., Davies, P., & Richman, B. (1971). *Word frequency book*. New York: American Heritage Publication Co.

Chomsky, C. (1970). Reading, writing, and phonology. *Harvard Educational Review, 40*, 287-309.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Cook, L. (1981). Misspelling analysis in dyslexia: Observation of developmental strategy shifts. *Bulletin of the Orton Society, 31*, 123-134.

Derwing, B. L. (1976). Morpheme recognition and the learning of rules for derivational morphology. *Canadian Journal of Linguistics, 21*, 38-66.

Derwing, B. L., & Baker, W. J. (1979). Recent research on the acquisition of English morphology. In P. Fletcher & M. Garman (Eds.), *Language acquisition*. Cambridge: Cambridge University Press.

Fischer, F. W., Shankweiler, D., & Liberman, I. Y. (1985). Spelling proficiency and sensitivity to word structure. *Journal of Memory and Language, 24*, 282-295.

Freyd, P., & Baron, J. (1982). Individual differences in acquisition of derivational morphology. *Journal of Verbal Learning and Verbal Behavior, 21*, 282-295.

Frith, U. (1980). Unexpected spelling problems. In U. Frith (Ed.), *Cognitive processes in spelling*. New York: Academic Press.

Goyen, J. D., & Martin, M. (1977). The relation of spelling errors to cognitive variables and word type. *British Journal of Educational Psychology, 47*, 268-273.

Hanna, P. R., Hodges, R. E., & Hanna, J. S. (1971). *Spelling: Structure and strategies*. Boston: Houghton Mifflin.

Hodges, R. E., & Rudorf, E. H. (1966). Searching linguistics for cues for the teaching of spelling. In T. D. Horn (Ed.), *Research in handwriting and spelling*. Champaign, IL: National Council of English Teachers.

Holmes, D. L., & Peper, R. J. (1977). An evaluation of the use of spelling error analysis in the diagnosis of reading disabilities. *Child Development, 48*, 1708-1711.

Jastak, J. F., & Jastak, S. (1978). *The Wide Range Achievement Test*. Wilmington, DE: Jastak Associates, Inc.

Jorm, A. F. (1981). Children with reading and spelling retardation: Functioning of whole-word and correspondence-rule mechanisms. *Journal of Child Psychology and Psychiatry, 22*, 171-178.

Klima, E. S. (1972). How alphabets might reflect language. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, MA: MIT Press.

Larsen, S. C., & Hammill, D. D. (1976). *Test of Written Spelling*. Austin, Texas: Pro-Ed.

Liberman, I. Y. (1982). A language-oriented view of reading and its disabilities. In H. Myklebust (Ed.), *Progress in learning disabilities*. New York: Grune and Stratton.

Marino, J. L. (1979). *Children's use of phonetic, graphemic, and morphophonemic cues in a spelling task*. Educational Resources Information Center, Document #ED188235.

Moats, L. C. (1983). A comparison of the spelling errors of older dyslexics and second-grade normal children. *Annals of Dyslexia, 33*, 121-139.

Nelson, H. (1980). Analysis of spelling errors in normal and dyslexic children. In U. Frith (Ed.), *Cognitive processes in spelling*. New York: Academic Press.

Rubin, H. (1984). *An investigation of the development of morphological knowledge and its relationship to early spelling ability*. Unpublished doctoral dissertation, University of Connecticut.

Schwartz, S., & Doehring, D. G. (1977). A developmental study of children's ability to acquire knowledge of spelling patterns. *Developmental Psychology, 13*, 419-420.

Selby, S. (1972). The development of morphological rules in children. *British Journal of Educational Psychology, 42*, 293-299.

Smith, P. T. & Sterling, C. M. (1982). Factors affecting the perceived morphemic structure of written words. *Journal of Verbal Learning and Verbal Behavior, 21*, 704-721.

Sterling, C. M. (1983). Spelling errors in context. *British Journal of Psychology, 74*, 353-364.

Sweeney, J. E., & Rourke, B. P. (1978). Neuropsychological significance of phonetically accurate and phonetically inaccurate spelling errors in younger and older retarded spellers. *Brain and Language, 6*, 212-225.

Templeton, S. (1980). Spelling, phonology, and the older student. In E. H. Henderson & J. W. Beers (Eds.), *Developmental and cognitive aspects of learning to spell: A reflection of word knowledge*. Newark, DE: International Reading Association.

Templeton, S., & Scarborough-Franks, L. (1985). The spelling's the thing: Knowledge of derivational morphology in orthography and phonology among older students. *Applied Psycholinguistics, 6*, 371-390.

Whiting, S. A., & Jarrico, S. (1980). Spelling patterns of normal readers. *Journal of Learning Disabilities, 13*, 45-47.

Wiig, E. H., Semel, E. M., & Crouse, M. A. B. (1973). The use of English morphology by high-risk and learning-disabled children. *Journal of Learning Disabilities*, 7, 59-67.

Venezky, R. L. (1970). *The structure of English orthography*. The Hague: Mouton.

Venezky, R. L. & Weir, R. H. (1966). *A study of selected spelling-to-sound correspondences*. Stanford, CA: Stanford University.

# THE DEVELOPMENT OF MORPHOLOGICAL KNOWLEDGE IN RELATION TO EARLY SPELLING ABILITY

**Hyla Rubin**†

**Abstract.** *This study assessed the morphological knowledge of kindergarteners and first graders in relation to their early spelling ability. Morphological knowledge was investigated because, in order to spell, children need to understand that words are composed of morphemes and phonemes, and because poor spellers have particular difficulty with inflected forms of words. Kindergarteners and first graders were grouped by their implicit understanding of morphology and were given tests of dictated spelling and morphological analysis. First graders with poor morphological knowledge omitted more inflectional morphemes in spelling and were less able to identify base morphemes in spoken words than kindergarteners and first graders with higher levels of implicit morphological knowledge. The results demonstrate the importance of morphological knowledge in the development of spelling proficiency.*

## INTRODUCTION

Children who demonstrate learning problems characteristically make errors when reading and spelling inflected and derived forms of words. They tend to omit and substitute inflectional markers and to substitute base words for derived words, or one derived form of a word for another. Although these errors are frequently documented in clinical case reports, there is little experimental research concerning morphemic errors in written language. The studies that do exist demonstrate that children with learning problems make more of these errors in spelling than other children (Anderson, 1982; Moran, 1981). However, possible reasons for the occurrence of these errors have not been addressed.

The basis for such errors in spelling might fall into one of two categories. On the one hand, they might represent part of a general tendency to misspell words. If this is the case, omissions of inflectional endings, for example, might be but one instance of a more pervasive pattern of final consonant omissions. On the other hand, they might reflect an underlying deficit in morphological

---

† University of Toronto
*Acknowledgment.* This work was supported by NICHD grant HD-01994 to Haskins Laboratories and by a dissertation fellowship from the University of Connecticut Research Foundation. The author wishes to thank Isabelle Y. Liberman, Ignatius G. Mattingly, Donald P. Shankweiler, Leonard Katz, and Robert Katz for their invaluable assistance throughout this project.

knowledge. If that is the case, children who make such errors in spelling would be expected to perform poorly in their attempts to use morphological rules in spoken language or to analyze the internal structure of words.

Although the relationship between morphological knowledge and spelling ability has not been examined directly, there is good reason to anticipate that children who make morphemic errors in spelling are indeed deficient in their underlying morphological skills. Several studies have demonstrated that children with reading problems have difficulty applying morphological rules to new words (Brittain, 1970; Doehring, Trites, Patel, & Fiedorowicz, 1981; Vogel, 1975, 1983; Wiig, Semel, & Crouse, 1973). In all of these studies, morphological knowledge has been assessed by an elicited spoken language task that requires the application of basic inflectional and derivational rules of morphology to nonsense base words (Berko, 1958; Berry, 1966). This method is used in order to determine that children are actually applying the morphological rules that they have mastered and are not just producing memorized vocabulary items. It has been found that normally developing children master these rules between the ages of four and seven (Brown, 1973; deVilliers & deVilliers, 1973; Selby, 1972; Templin, 1957). In contrast, children with learning problems develop morphological knowledge more slowly, although they are found to follow the same sequence in their rule acquisition.

Although it has been demonstrated that children grouped by reading ability differ significantly in their use of inflectional morphemes, as measured by the Berko procedure, research has not yet examined whether morpheme use is directly related to other linguistic skills or why these relationships might exist. Since children with learning problems show a strong tendency to make morphemic errors in spelling as well as in reading, it is of particular interest to determine if there is a relationship between morphological knowledge and spelling ability. Since the English orthography is morphophonemic, like the spoken language it represents (Liberman, Liberman, Mattingly, & Shankweiler, 1980), spelling requires that the child understand that words are made up of morphemes, which, in turn, are made up of phonemes. Studies of spelling ability of college students indicate that poor spellers fail most dramatically on those words that require sensitivity to morphophonemic structure (Fischer, 1980; Hanson, Shankweiler, & Fischer, 1983). In addition, examination of the spontaneous writing samples of learning disabled children and adults documents incorrect usage of both inflectional and derivational morphemes (Anderson, 1982; Liberman, Rubin, Duques, & Carlisle, 1985; Moran, 1981). Based on this evidence, a strong relationship between the ability to use morphemes correctly in spoken and written language would be expected since morpheme use in either case would depend on the development of morphological rules and access to them in the lexicon. It would also be expected that morpheme use would depend, at the very least, on an implicit understanding of morphophonemic structure. However, the explicit understanding that words are made up of morphemes, which, in turn, are made up of phonemes, would clearly differentiate the proficient from the disabled writers.

Previous research studies have demonstrated that the ability to analyze the internal structure of words explicitly is a critical component in learning to read (Blachman, 1983; Fox & Routh, 1980; Liberman, Shankweiler, Fischer, & Carter, 1974; Lundberg, Olofsson, & Wall, 1980; Treiman & Baron, 1981) and in learning to spell (Liberman et al., 1985; Perin, 1983; Zifcak, 1981). In the reading studies, the ability to analyze spoken words into syllabic and phonemic segments has been found to be highly related to letter naming and word recognition performance in kindergarten, first- and second-grade children. In the spelling studies, phonemic segmentation ability has been found to be significantly related to dictated spelling performance in kindergarteners (Liberman et al., 1985), first graders (Zifcak, 1981), and adolescents (Perin, 1983).

Research into the structural analysis of spoken words and its relationship to reading and spelling abilities has yielded valuable diagnostic and instructional information thus far. It is clear that children with reading and spelling problems are less able than their normally achieving peers to analyze spoken words into their constituent phonemes. However, many questions about this relationship remain unanswered. To begin with, the ability to analyze spoken words into their constituent morphemes has been barely examined. Since the English orthography, like the spoken language it represents, is morphophonemic, we need to investigate the ability to analyze the internal structure of words as a function of both morphemic and phonemic structure.

Recent studies have begun to examine the explicit understanding of morphophonemic structure in children. Derwing and Baker (1977, 1979) have investigated the development of morpheme identification ability in children in grades 3 through college. They provided the children with word pairs that were varied for semantic and phonetic similarity, such as *teach-teacher, slip-slipper, cup-cupboard*, and *moon-month*. The children were required to read each pair and indicate if one word "came from the other," using a 5-point scale to specify the degree of relatedness. Performance correlated with age and degree of semantic and phonetic relationship between the paired words. The authors concluded that morpheme recognition ability may develop as much through instructional experience as through language acquisition and suggested that it would be difficult to sort out the contributions of these two sources of linguistic knowledge.

Although this research into the explicit analysis of morphemic structure is provocative, similar studies have not been conducted with children who demonstrate learning problems or with children below third grade. It would be expected that if younger children were deficient in morpheme use, which would reflect their implicit awareness of morphological structure, they would also be deficient in their ability to recognize base morphemes within two-morpheme words, or their explicit awareness of morphological structure. If these abilities were found to be related to each other and to morpheme use in early spelling, it would be possible to demonstrate the necessity of helping young children develop sensitivity to morphemic structure through direct instruction.

Therefore, the present study was designed to examine the relationship between implicit awareness of morphemic structure, as measured by the ability to apply morphological rules to new words, and explicit awareness of morphemic structure, as measured by the ability to identify base words within two-morpheme words. Furthermore, the relationship between performance on the spoken language tasks and the ability to represent base morphemes and inflectional morphemes in beginning attempts at spelling was investigated.

Although previous studies that document morphemic errors in spelling analyzed spontaneous writing samples, it was not considered reasonable to elicit writing samples in the present study since the children tested were only in kindergarten and first grade. However, it was important to select children of this age for several reasons. First of all, it was expected that they would demonstrate sufficient variability in their levels of implicit and explicit awareness of morphological structure of spoken words to enable us to learn more about the course of this development. Secondly, previous studies of invented spelling (Read, 1971, 1975) have demonstrated that by age five many children are able to analyze words into their constituent phonemes and use their knowledge of letter names to "invent" written representations of the spoken words. By scoring for the number of morphemes represented in writing rather than for correctness of spelling, it seemed reasonable to use a dictated spelling task as an early indication of the ability to represent inflectional morphemes in written form. In this way, both spoken and written language measures of the morphological knowledge of young children could be obtained. Finally, this information could be used in future research to predict the course of morphemic development in the written language of children and adults.

## Method

### Subjects

The subjects were children selected from four kindergarten classes and four first-grade classes in a suburban Connecticut public school. The children eligible for testing were all those for whom parental permission was obtained. The available 128 children (59 kindergarteners and 69 first graders) demonstrated adequate vision and hearing and were judged to have normal intelligence by their classroom teachers and the school psychologist. During a one-week period, they were individually given the *Berry-Talbott Test of Language* (Berry, 1966), a measure of elicited morpheme production in spoken language. This test required them to apply basic inflectional and derivational rules of morphology to nonsense base words by completing spoken sentences when shown illustrative pictures.

Four groups were formed by selecting those children from each grade who scored within the highest and lowest thirds of the distribution of scores on the Berry-Talbott Test of Language. The children from the highest third of the kindergarten and first-grade distributions will be referred to as the high kindergarteners and high first graders. Similarly, the subjects from the lowest third of the kindergarten and first grade distributions will be referred to as the low kindergarteners and low first graders. The mean age and test scores for each group are summarized in Table 1.

**Table 1**

**Berry-Talbott Test of Language: Grouped Mean Score (and Standard Deviation) for Kindergarteners and First Graders**

| | Low Kindergarten | High Kindergarten | Low First Grade | High First Grade |
|---|---|---|---|---|
| n | 21 | 19 | 22 | 24 |
| Berry-Talbott | 10.8 | 24.7 | 14.1 | 28.0 |
| | (3.3) | (2.5) | (4.1) | (3.3) |
| Age (years-months) | 5-5 | 5-5 | 6-5 | 6-5 |

To determine if the children differed in their performance on the Berry-Talbott Test, an analysis of variance was conducted. The analysis revealed a significant main effect of group (high, low), $F(1,82) = 347.16, MS\epsilon = 11.83, p < .001$, and grade (kindergarten, first), $F(1,82) = 19.92, MSe = 11.83, p < .001$. There was no interaction between group and grade. Furthermore, comparison tests revealed significant differences among the groups: the high first graders performed better than the high kindergarteners, $t(41) = 3.58, p < .001$; the low first graders performed better than the low kindergarteners, $t(41) = 2.86, p < .007$; and the high kindergarteners performed better than the low first graders, $t(39) = 9.49, p < .001$.

## Materials and Specific Procedures

1) *Experimental Spelling Test.* This measure was designed to assess the children's representation of base and inflectional morphemes in the early stages of their experience with written language. It contained 31 words that were considered to be part of the average kindergartener's spoken vocabulary. Twenty-one words were organized according to morphemic structure (one or two morphemes) and type of final consonant cluster (nasal or non-nasal). Three experimental words were given in each of the following categories: (1) 2-morpheme words ending in *md (hummed, jammed, dimmed)*, (2) 1-morpheme words ending in *nd (wind, band, kind)*, (3) 2-morpheme words ending in *nd (pinned, canned, lined)*, (4) 1-morpheme words ending in *nt (tent, pant, hint)*, (5) 2-morpheme words ending in *nt (bent, can't, don't)*, (6) 1-morpheme words ending in *st (list, dust, nest)*, and (7) 2-morpheme words ending in *st (kissed, fussed, messed)*. Ten words were used as fillers to reduce the possible priming effects of the experimental words. Five of the fillers were one-morpheme words *(winter, candy, dinner, money*, and *wise)* and five were two-morpheme words *(hunter, windy, winner, funny*, and *pies)*. The experimental and filler words were randomized and each word was dictated, then used in a meaningful sentence and repeated. The children were instructed to write each word on a pre-numbered response form.

(2) *Experimental Morpheme Analysis Test.* This measure was designed to assess the ability to analyze a spoken word into its constituent morphemes by requiring each child to identify base morphemes within words. This task consisted of the same 31 words that were used for spelling. The child was asked questions such as "Is there a smaller word in *dust* that means something like *dust*?" or "Is there a smaller word in *kissed* that means something like *kissed*?" for each of the words. For one-morpheme words (such as *dust, pant*, and *wind)*, the child was supposed to respond "No." For two-morpheme words (such as *fussed, can't*, and *pinned)*, the child was supposed to respond "Yes" and supply the base word.

These procedures were demonstrated in six training trials in the following manner. First, the child listened to each question and responded spontaneously. If the response was incorrect, the examiner repeated the question, provided the correct response along with a brief explanation, and asked the question again. This procedure was repeated once if needed. Words that contained smaller words that were not related to the stimulus word (such as *pillow* and *sink*) were included in the training trials and required "no" responses. On the test trials, no demonstrations or feedback were given.

## General Procedures

The 86 children in the four groups were tested further to determine the relationship of their morpheme use in spoken language to their morpheme use in spelling and to their explicit morpheme analysis ability. During the one-week period following administration of the *Berry-Talbott Test of Language* (1966), each of the four groups of children was given the dictated experimental spelling test in a half-hour group session. During the following three-week period, each child was given the experimental morpheme analysis task and a letter naming task in an individual testing session of approximately 20 minutes. To insure consistent presentation of the stimuli, all of the test items were presented on tape.

## Results

### Implicit Morphological Knowledge and Spelling Ability

Letter naming scores were tabulated and showed that all but the low kindergarten children could name over 90% of the letters of the alphabet, a skill needed for invented spellings.

For each child, the percentage of written words with final consonant omissions was also tabulated. The high first graders omitted final consonants from 3% of the words, the high kindergarteners from 10% of the words, and the low first graders from 17% of the words. (Since low kindergarteners were not able to name the letters of the alphabet, their spelling results will not be discussed.) To determine if the groups differed in their tendency to omit final consonants, an analysis of variance was conducted with two between-groups factors (implicit morphological knowledge in spoken language, grade level). The analysis revealed a significant main effect of implicit morphological knowledge, $F(1,82) = 4.25, MSe = 5.97, p < .043$, and a significant interaction between morphological knowledge and grade level, $F(2,82) = 12.63, MSe = 5.97, p < .001$.

These results suggest that the ability to represent final consonants in written language is significantly related to morphological knowledge in spoken language and is not significantly related to grade level independent of linguistic ability. In other words, the low first graders omitted more final consonants than did either the high first graders or the high kindergarteners.

When the data are examined as a function of both morphemic and phonemic structure, they indicate that in omitting final consonants in their spelling, children tend not to be influenced by the phonemic structure of the words. It was found that the percentage of error on words ending in nasal and non-nasal consonant clusters was roughly the same–8% and 7%, respectively. In contrast, there was a striking effect of morphemic structure. Whereas children omitted final consonants from only 4% of one-morpheme words, they omitted final consonants from 11% of two-morpheme words, a difference that was highly significant, $t(85) = 5.84, p < .001$. It is clear from these results that final consonants were omitted more often from two-morpheme than from one-morpheme words, and that it was the morphologically less knowledgeable first graders who were omitting those inflectional morphemes.

### Implicit and Explicit Levels of Morphological Knowledge

In the morpheme analysis task, a two-morpheme word (such as *pinned*) was scored as correct if the child (1) responded "Yes" and supplied the correct base form of the word (*pin*), and (2) responded "No" to a phonemically similar one-morpheme word (wind). (The *md* words [*hummed, jammed, dimmed*] were excluded from this scoring system because there are no one-morpheme words in English that end in *md*.) The two-pronged scoring system was necessary to counter possible effects of response bias. Without such a system, indiscriminate "no" responses would result in higher scores than indiscriminate "yes" responses, since "yes" responses had to be accompanied by the correct base word and "no" responses had no such control. By pairing words with similar phonemic structure and contrasting morphemic structure, one could be certain that "correct" responses validly represented sensitivity to morphemic structure and not inflation due to response bias.

Using this scoring system, the percentage of correctly analyzed word pairs was tabulated for each child. Both high first graders and high kindergarteners analyzed 48% of the pairs correctly, low first graders 24%, and low kindergarteners 3%. The correlation between the number of pairs a child analyzed correctly and morpheme use in spoken language proved to be significant, $r(84) = .63, p < .001$, indicating a strong relationship between implicit and explicit levels of morphological awareness.

To determine if the groups of children differed in their ability to identify base morphemes in pairs of words that differed in morphemic complexity, an analysis of variance was conducted with two between-groups factors (implicit morphological knowledge in spoken language, grade level). The analysis revealed a significant main effect of implicit morphological knowledge, $F(1,82) = 49.11, MSe = .05, p < .001$, and grade, $F(1,82) = 5.80, MSe = .05, p < .019$. Moreover, the interaction between morphological knowledge and grade level was significant, $F(2,82) = 4.31, MSe = .05, p < .042$. In other words, the high kindergarteners and high first graders performed equally well.

These results show that implicit morphological knowledge in spoken language (as assessed by the *Berry-Talbott Test*) is a more important discriminator of explicit morphological knowledge than is grade level. Implicit morphological awareness in spoken language accounted for 34% of the total variance in explicit morphological awareness, whereas grade level accounted for only 4%, and the interaction between group and grade for 3%.

What is particularly notable about these results is that children with high levels of implicit morphological knowledge in the elicited spoken language task performed equally well on the explicit analysis task regardless of grade level differences. Therefore, the ability to analyze morphemic structure explicitly, at least as measured by this task and at this point in development, seems to be more highly related to implicit morphological knowledge in spoken language than to grade level factors such as age and amount of instructional experience.

## Discussion

The purpose of this study was to investigate the development of morphological knowledge and its relationship to early spelling ability in kindergarten and first-grade children. Two levels of morphological knowledge were examined, since previous research has suggested that children need to understand morphophonemic structure implicitly and explicitly in order to spell. Although previous studies had shown that written language proficiency requires an explicit understanding of morphophonemic structure, the ability of young children to analyze the internal structure of words had been examined at the phonemic but not at the morphemic level of language.

It was found, in accordance with previous studies of normal language acquisition, that children in kindergarten and first grade are still developing implicit morphological knowledge (as measured by the *Berry-Talbott*), and that they use certain morphological rules before others. Notably, in view of the large number of past tense items in the stimuli that were used to assess spelling and explicit analysis abilities, most of the kindergarteners and first graders in this study successfully applied the morphological rules for regular past tense (in the nonsense words: *trommed, flitched, linged*, and *bazinged*).

In addition, it was found that implicit morphological knowledge does not develop solely as a function of factors associated with grade level. This was seen by the fact that some kindergarteners (the high group) performed significantly better than some first graders (the low group).

However, the role of factors associated with grade level cannot be disregarded either, since high first graders performed significantly better than high kindergarteners, and low first graders performed significantly better than low kindergarteners. What is clear from these results is that kindergarteners and first graders vary greatly in their implicit knowledge of the morphology and that this variability affects their early spelling ability.

In fact, implicit morphological knowledge had a more significant effect than grade level on the tendency of young children to omit inflectional morphemes in spelling. This was seen by the fact that low first graders made relatively more of these errors than either high first graders or high kindergarteners. Furthermore, the poorly developed implicit morphological knowledge of the low first graders correlated highly with their poor performance on the morphemic analysis task.

Considering previous research on phonemic analysis, it was enlightening to examine the types of errors made by the low kindergarteners and low first graders when they attempted to analyze the morphemic structure of spoken words explicitly. It was found that many of these children could manipulate phonemic segments without understanding morphemic structure. For example, in response to the questions "Is there a smaller word in *kind* that means something like *kind*?" and "Is there a smaller word in *dust* that means something like *dust*?", they often responded "*Yes, kin*" or "*Yes, tind* or "*Yes, dus*" or *Yes, tust*." This finding highlights the importance of examining the ability to explicitly analyze the morphemic structure as well as the phonemic structure of words.

Looking more closely at the results of the explicit morphemic analysis task, the fact that the high kindergarteners and high first graders performed with identical proficiency, despite their different amounts of instructional experience, raises an interesting question. Since the high first graders demonstrated a significantly higher level of implicit morphological knowledge than the high kindergarteners, it seems curious at first that these two groups demonstrated identical levels of explicit morphological knowledge. Apparently, the high first graders would have had to show a greater superiority in implicit morphological awareness over the high kindergarteners in order to demonstrate a more sophisticated level of explicit awareness. In addition, the explicit analysis task may not have been sensitive enough to detect differences between the two high groups. What seems clear is that the ability to analyze the morphophonemic structure of a word is to some degree independent of instructional experience at this age level, since high kindergarteners performed significantly better than low first graders. Since it is difficult to sort out the roles of linguistic ability and instructional experience at higher age levels, it is particularly helpful to begin to sort out these contributions for young children. By doing so, we can begin to develop more sensitive diagnostic measures to predict later language learning deficits and to design instructional procedures that will address the morphophonemic aspects of learning to read and spell.

The present study demonstrates that children in both kindergarten and first grade vary considerably in their implicit and explicit knowledge of the morphology and that this variability affects their early attempts to represent base and inflectional morphemes in writing. It is clear from the obtained results that children who demonstrate weak implicit knowledge of morphological rules are also deficient in their ability to explicitly analyze the internal morphemic structure of words and to use inflectional morphemes in writing. Therefore, the greater tendency of the low first graders to omit inflectional morphemes in writing seems to reflect a deficiency in morphological knowledge, rather than just a general spelling problem.

It is notable that, even though most of the children demonstrate their implicit knowledge of the past tense rule on the Berry-Talbott Test, only the children in the high groups show some degree of proficiency when explicitly analyzing the internal morphemic structure of past tense words. In contrast, the children in the low groups are relatively unable to analyze the internal morphemic structure of the past tense words, and omit relatively more past tense inflectional morphemes in writing. Yet they too were able to use the morphological rule for past tense on the Berry-Talbott Test. At least for the low first graders, this pattern of performance suggests that it is their lack of explicit awareness of morphemic structure that should cause us the most concern. Although these children demonstrate some ability to manipulate phonemic structure, based on the errors they made on the morpheme analysis task, they do not seem to understand that inflected words are composed of groups of phonemes that form morphemes. Therefore, it seems probable that their lack of explicit understanding of morphophonemic structure, in conjunction with their generally weak implicit knowledge of the morphology, account in large measure for the morphemic errors they make in their early spelling attempts.

It seems clear, then, that even at the primary level, if children are to be good spellers, it is not enough for them to understand that words are made up of phonemic segments. Research into the spelling and written expression performance of older children and adults with learning problems demonstrates that errors on inflected and derived forms of words are a major characteristic of their written products. The results of this study suggest that the basis for such errors may be an underlying deficiency at the implicit level, and especially at the explicit level, of morphological knowledge. Therefore, it is of critical importance that we assess the morphological knowledge of young children so that we may identify those who are at risk for learning problems and help them to develop the sensitivity to morphophonemic structure that they need to become proficient written language users.

In order to best help these children, it seems necessary to teach them to use grammatical morphemes correctly in their spoken language if they are to become competent in spelling inflected and derived forms of words. In addition, the present results suggest that it is critical to teach these children to become *explicitly* aware of the structure of their spoken language productions. It is this explicit awareness of their language that should help children to apprehend the internal structure of the new words that they are required to read and spell. Written language instruction should focus on the development of structural analysis skills at both the morphemic and phonemic levels. It is clear that children should be taught that words (whether they are spoken, read, or spelled) are composed of morphemes, which, in turn, are composed of phonemes.

In conclusion, this study represents a first step in the examination of morphological knowledge in the spoken language of young children as it relates to their ability to represent morphemes in writing and their ability to analyze the internal morphemic structure of words. Since this is a new area of investigation, it is anticipated that these results will stimulate the development of other research studies. In the future, we need to conduct similar studies with learning disabled children, adolescents, and adults in an effort to account for the morphemic errors they make in reading and written expression. In this way, we can begin to document deficiencies in sensitivity to morphophonemic structure in these groups. It is hoped that studies of this type will result in improved diagnostic and instructional procedures for children and adults with language-learning disabilities.

## References

Anderson, P. L. (1982). A preliminary study of syntax in the written expression of learning disabled children. *Journal of Learning Disabilities, 15*, 359-362.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*, 150-177.

Berry, M. (1966). *Berry-Talbott Language Test: Comprehension of Grammar*. Rockford, Ill.

Blachman, B. (1983). Are we assessing the linguistic factors critical in early reading? *Annals of Dyslexia, 33*, 91-109.

Brittain, M. (1970). Inflectional performance and early reading achievement. *Reading Research Quarterly, 6*, 34-48.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

deVilliers, J., & deVilliers, P. (1973). A cross-sectional study of the acquisition of grammatical morphemes. *Journal of Psycholinguistic Research, 2*, 267-278.

Derwing, B., & Baker, W. (1977). The psychological basis for morphological rules. In J. McNamara (Ed.), *Language learning and thought*. New York: Academic Press.

Derwing, B., & Baker, W. (1979). Recent research on the acquisition of English morphology. In P. Fletcher & M. Garman (Eds.), *Language acquisition*. Cambridge: Cambridge University Press.

Doehring, D., Trites, R., Patel, P., & Fiedorowicz, C. (1981). *Reading disabilities: The interaction of reading, language, and neuropsychological deficits*. New York: Academic Press.

Fischer, F. W. (1980). *Spelling proficiency and sensitivity to linguistic structure*. Unpublished doctoral dissertation, University of Connecticut.

Fox, B., & Routh, D. (1980). Phoneme analysis and severe reading disability in children. *Journal of Psycholinguistic Research, 9*, 115-119.

Hanson, V. L., Shankweiler, D., & Fischer, F. W. (1983). Determinants of spelling ability in deaf and hearing adults: Access to linguistic structure. *Cognition, 14*, 323-344.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. Kavanaugh & R. Venezky (Eds.), *Orthography, reading, and dyslexia*. Baltimore, MD: University Park Press.

Liberman, I. Y., Rubin, H., Duques, S., & Carlisle, J. (1985). Linguistic abilities and spelling proficiency in kindergarteners and adult poor spellers. In D. B. Gray & J. F. Kavanaugh, *Biobehavioral measures of dyslexia*. Parkton, MD: York Press.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phoneme awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159-173.

Moran, M. R. (1981). *A comparison of formal features of written language of learning disabled, low-achieving and achieving secondary students* (Research Report No. 34). Lawrence, Kansas: University of Kansas Institute for Research in Learning Disabilities.

Perin, D. (1983). Phonemic segmentation and spelling. *British Journal of Psychology, 74*, 129-144.

Read, C. (1971). Pre-school children's knowledge of English phonology. *Harvard Educational Review, 41*, 1-34.

Read, C. (1975). *Children's categorizations of speech sounds in English*. NCTE Committee on Research Report No. 17 (Urbana, Illinois), Eric Document Reproduction Service No. Ed 112 426.

Selby, S. (1972). The development of morphological rules in children. *British Journal of Educational Psychology, 42*, 293-299.

Templin, M. (1957). *Certain language skills in children: Their development and interrelationships.* Minneapolis: University of Minnesota Press.

Treiman, R., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. In G. E. MacKinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 3). New York: Academic Press.

Vogel, S. (1975). *Syntactic abilities in normal and dyslexic children.* Baltimore: University Park Press.

Vogel, S. (1983). A qualitative analysis of morphological development in learning disabled and achieving children. *Journal of Learning Disabilities, 16*, 416-420.

Wiig, E., Semel, E., & Crouse, M. (1973). The use of morphology by high-risk and learning disabled children. *Journal of Learning Disabilities, 6*, 457-465.

Zifcak, M. (1981). Phonological awareness and reading acquisition. *Contemporary Educational Psychology, 6*, 117-126.

# THE CROSSWORD PUZZLE PARADIGM: THE EFFECTIVENESS OF DIFFERENT WORD FRAGMENTS AS CUES FOR THE RETRIEVAL OF WORDS*

**Naomi Goldblum**† and **Ram Frost**‡

**Abstract.** *We investigated the internal structure of words in the mental lexicon by using a crossword puzzle paradigm. In two experiments, subjects were presented with word fragments along with a semantic cue, and were asked to retrieve the whole word that contained the presented fragment and was compatible with the semantic information. In Experiment 1, we found that any cluster of adjacent three letters facilitated retrieval better than dispersed letters. Moreover, syllabic clusters had greater facilitative effect than nonsyllabic pronounceable clusters, or nonpronounceable clusters. In Experiment 2, we found that syllabic units facilitated retrieval more than morphemic units. The results are interpreted as evidence for the existence of lexical subunits that are larger than the letter but smaller than the word, and that are organized according to phonologic principles. An interactive model for solving crossword puzzles is proposed.*

## INTRODUCTION

This paper is concerned with the following question: Does the mental lexicon contain units smaller than the whole word but larger than the individual letter, and if so, what kind of units are they? The previous answers to these questions seem to be modality-specific. There is wide agreement that syllabic units play an important role in auditory word perception (e.g., Kahn, 1976; Mehler, Dommergues, Frauenfelder, & Segui, 1981; Segui, 1984). In research on visual word perception, on the other hand, there is conflicting evidence as to what the subword units might be, and whether or not the visually presented stimuli undergo phonologic as well as visual processing.

---

* *Memory & Cognition*, in press.

† Department of Psychology, The Hebrew University, Israel

‡ Also Department of Psychology, The Hebrew University, Israel

*Acknowledgment.* The authors wish to thank Len Katz, Bruno Repp, Laurie Feldman, Susan Brady, Catherine Browman, Benny Shanon, Itzhak Schlesinger, and two anonymous reviewers for useful comments on earlier drafts of this paper. This work was supported in part by National Institute of Child Health and Human Development Grant HD-09144 to Haskins Laboratories. The study is based on a master's dissertation presented by the first author to the Department of Psychology, the Hebrew University.

Spoehr and Smith (1975) have shown that a vocalic center group (VCG) is more easily percieved than a similar cluster of letters not containing a vowel. Their use of the VCG is based on the work of Hansen and Rodgers (1965), who define a VCG as a cluster consisting of a vowel with a consonant or consonants on either side, where the whole cluster forms a pronounceable unit. AN, CAN, ANT, and CANT are examples of VCGs. They have also shown that one-syllable words are processed faster and more accurately than two-syllable words of the same number of letters (Spoehr & Smith, 1973; see also Spoehr, 1978). From these results Spoehr and her colleagues concluded that words are represented in the lexicon according to their syllabic structure.

In contrast to the phonological division suggested by Spoehr and her associates, Murrell and Morton (1974) and Taft and Forster (1975) proposed a morphological division into units. According to this view, polymorphemic words are stored in the lexicon in a morphologically decomposed fashion: the root and the information about prefixes and inflections. Thus, in the process of word recognition, the reader strips the prefixes, and accesses the morphological root first. A different division of written words into units was suggested by Taft (1979). Taft defined the minimal lexical unit as a Basic Orthographic Syllabic Structure (BOSS). The BOSS is formed by adding as many consonants as possible to the first vowel in the first syllable, without violating the orthotactic rules of English. Thus, the BOSS is a unit consisting of as many consonants as can legally be found together with one vowel, at the beginning or the end of a word. According to this view, in order to access a multimorphemic word in the mental lexicon, one first accesses its BOSS unit. In a series of experiments designed to investigate Taft's hypothesis, Lima and Pollatsek (1983) found no difference between the facilitative effect of syllables and BOSS units. They demonstrated, however, that either of these units was better than an arbitrary unit in priming a word of which they were a constituent. When a syllabic unit was also a morphemic unit of the word, then it was more facilitative than a syllabic unit that did not constitute a morpheme of the word.

This inconsistency of results is puzzling but may perhaps be attributed to task characteristics. All of the above studies concern visual word perception, and most of them use the lexical decision paradigm. Usually, in the experiments described above, words that are parsed into units according to phonologic or orthographic principles are presented visually to the subject. Here, the speed and accuracy of lexical decisions to such parsed words is assumed to reflect the naturalness of these units. It is assumed that if lexical search is facilitated by a particular division of a word, then this division actually reflects important characteristics of the representation of this word in the internal lexicon. However, it has been recently suggested that lexical decisions, in many cases, do not involve more than superficial lexical access (Balota & Chumbley, 1984). Since all that is needed for lexical decision is a judgment concerning the probability that the letter string is a valid word, it is possible that, for at least some words, the decision is based on a fast judgment concerning the familiarity of the letter string. In such case, the decision stage occurs prior to any deep analysis of meaning and morphemic structure. This suggestion is described in a two-stage model of lexical decision performance (Balota & Chumbley, 1984). According to this model, very familiar and very unfamiliar letter strings are processed superficially without lexical access. The letter string will undergo deeper processing that involves decomposition into units only when a fast decision concerning its familiarity cannot be reached. Consequently, in a lexical decision task where a whole word is presented to the subject, a division of the word into subunits may, in many cases, be irrelevant to the task. If this is the case, then the structure of the internal lexicon may not be accurately reflected by performance in lexical decision experiments.

A retrieval task, on the other hand, can avoid the artifacts of the lexical decision process. If only a fragment of a word is presented, and the subject is asked to retrieve the whole word containing this fragment, the extent to which a particular fragment facilitates retrieval may reflect the functional role of this fragment in the lexicon.

An example of such cue-facilitated retrieval is the process that occurs in the solving of crossword puzzles. When part of the word is filled in, the solver has two cues for the retrieval of the word: the filled-in letters in their appropriate places and the "definition," which is generally a synonym or some other associative term. When the solver cannot come up with the correct answer, he or she tries to fill in more letters by finding adjacent words. The solver usually chooses the position to be filled next, according to his or her intuition about the relative facilitatory effect of the positions that are still empty. This raises the followinq questions: What facilitates retrieval better, dispersed letters or letter clusters? Also, is there any difference among types of clusters? The study of the relative facilitatory effect of different types of word fragments may provide us, then, with useful clues about the structural representation of words in the mental lexicon. If words in the internal lexicon are actually organized in terms of subunits, it is more likely that people will make use of these subunits when they are presented with them and asked to retrieve the whole word, rather than simply make lexical decisions.

A number of experiments using the word-fragment paradigm indicate that with the number of letters controlled, all fragments are not equally effective for the retrieval of words. Horowitz, White, and Atwood (1968) presented subjects with lists of nine-letter words to memorize, and then tested whether the first, middle, or last three-letter fragment facilitated recall most. They found that the first fragment was most facilitative, followed by the last and middle fragment in that order of facilitation. However, Horowitz and his colleagues did not control the pronounceability of the fragments or whether they corresponded to syllables. This factor might have had some influence on the results. Since the middle fragment of a nine-letter word is less likely to be pronounceable than either of the end fragments, the position of the fragment may have been confounded with its pronounceability. Using a similar procedure, Dolinsky (1973) repeated this experiment with a control for the presence of syllables. After presenting his subjects with a list of words, recall was cued by presentation of syllabic and nonsyllabic fragments, at the beginning, middle, or final fragments of the word. Dolinsky found that the presence of a syllable had a significant facilitative effect on retrieval only in the middle fragments. When the cues were the begining or the final fragments, syllabic clusters did not facilitate recall better than nonsyllabic clusters. However, Dolinsky did not control for the pronounceability of the nonsyllable fragments and some of his nonsyllable controls were actually three letters of a four-letter syllable.

In the present study the word-fragment technique was used to investigate what sublexical word units, if any, exist in the internal lexicon when the letter's position within the word is controlled. It is possible (1) that individual letters in a word act separately and in parallel to activate directly the word of which they are constituents, or (2) that any group of consecutive letters in a word constitute a unit, or (3) that only very specific groups of consecutive letters have an activating effect greater than that of individual dispersed letters. If there are no middle-sized units in the lexicon, then all fragments of the same length should be equally helpful in retrieving a word. If letters grouped together are more effective in activating a word, then any group of consecutive letters should be a better retrieval cue than the same number of dispersed letters. If, however, there are specific groupings of letters that constitute units in the internal lexicon (e.g., syllables), then these specific groupings should be more effective cues for word retrieval than any other groupings of the same length.

## EXPERIMENT 1

Experiment 1 was designed to investigate whether letter clusters facilitate retrieval more than dispersed letters, and whether syllabic units are more facilitative than any other cluster of letters independently of their position in the word. To this end, syllabic units were compared with three types of fragments: pronounceable nonsyllabic clusters, unpronounceable clusters, and nonadjacent letters. To avoid the effect of length of cluster, all word fragments were composed of different combinations of three letters. For example, the target word "VINDICTIVE" was cued by the synonym "spiteful," together with one of the following four fragments:

1. _ _ _ DIC _ _ _ _ (syllable)
2. _ _ _ _ ICT _ _ _ (pronounceable nonsyllable)
3. _ _ NDI _ _ _ _ _ (unpronounceable cluster)
4. _ _ N _ I _ T _ _ _ (nonadjacent letters)

If there are no units larger than the individual letter in the internal lexicon, then any three letters of a word should be just as good a retrieval cue as any other three letters situated in similar positions within the word. If it is the clustering of the letters in itself that facilitates retrieval, then any cluster should be better than dispersed letters, without any difference between clusters of different types. If it is merely the pronounceability of the cluster that facilitates retrieval, then pronounceable clusters should be as facilitative as true syllables. If, however, syllables do constitute functional units in the internal lexicon, then a syllable should be more facilitative for the retrieval of the target word than any of the other fragments.

### Methods

**Subjects.** Sixty-four undergraduate students at the Hebrew University of Jerusalem participated in the experiment for course credit or for payment. All subjects were native English speakers.

**Stimuli and design.** The stimuli were 48 English words: 22 nouns, 8 verbs, and 18 adjectives. All the words had three syllables and were from seven to ten letters long. Their frequency, according to Kuçera and Francis (1967), ranged from 0 to 45, with a median of 10.5. There was no significant difference between the frequencies of the fragments of each type of cluster, according to the trigram frequency list presented by Underwood and Schulz (1960).

Four different types of fragments for each word were presented: A syllable, a pronounceable cluster that was not a syllable of this word, an unpronounceable cluster,[1] and three nonadjacent letters. Syllables were defined according to Webster's New World Dictionary of the American Language (1964). In those cases where the dictionary proposed two divisions, phonologic and orthographic, the phonologic division was used. All fragment types consisted of three letters; dashes were presented in place of all the missing letters. To eliminate the possibility that the number of vowels or consonants in the fragment might have some effect on retrieval, only fragments consisting of two consonants and one vowel were used. In order to ensure that the effect of the type of fragment was not confounded with the effect of the fragment's position, all the possible positions within the word were sampled. For the syllabic fragments, the first, the middle, and the last syllables were presented equally. In the isolated letters condition, half of the trials included either the first or the last letters of the word, and the remaining trials did not. The unpronounceable fragments were always in the middle of the word, as there are no words in which the first and the last fragments are unpronounceable, given the constraint that the fragment must contain a vowel. A semantic cue for the word, that is, a word or a phrase with approximately the same meaning as the target word, was presented in lowercase letters just above the letters-dashes configuration.

---

[1] By unpronounceable clusters we imply clusters that are phonotactically irregular in English.

Each word was presented with all four types of fragments, so as to serve as its own control. The subjects were divided into four groups. Each group was presented with only one of the four fragments of each word, in one of the possible fragment positions. Each group was presented with an equal number of words in each of the four fragment types. The different words in the different conditions were assigned to the four groups of subjects by means of a Latin square design, so that no subject saw a word more than once. The list of target words and fragments is presented in the Appendix.

**Procedure and apparatus.** The subjects were seated approximately 70 cm from a CRT screen in a semi-darkened room. Each stimulus appeared on the screen after the subject pressed a "start" button. The experimenter pressed a "finish" button when the correct answer was given by the subject, and only then was the stimulus removed from the screen. This procedure was deemed necessary because subjects often made incorrect spontaneous vocal responses. Consequently, a voice key for determining the exact reaction time could not have been used. However, in order to avoid an experimenter bias, the experimenter did not face the screen, and was not aware of the specific fragment condition in each trial. Rather, the experimenter was presented with a parallel list that contained all the correct responses, and pressed the "finish" button accordingly. If the subject gave an incorrect answer, he or she was told that it was incorrect and was allowed to guess again. If, however, the subject did not give the correct response in 30 seconds, the stimulus disappeared, reaction time (RT) was recorded as 30 seconds, and the trial was considered as a "no response" trial. Stimuli presentations and RT measurements were controlled by a PDP 11/23 computer. The subjects were presented with three practice trials before the test stimuli were presented.

### Results and Discussion

Reaction times in seconds and "no response" rates were calculated and averaged for the four experimental groups across the four fragment conditions. They are presented in Table 1.

The mean reaction times of each type of fragment were calculated across the different positions within the word. A one-way ANOVA revealed that the differences between the mean RTs to the different fragment types were significant, $F1(3,189) = 83.5, p < 0.001$ and $F2(3,141) = 29.5, p < 0.001, MinF' = 21.8, p < 0.05$.

## Table 1

### Mean Reaction Time in Seconds, Percent of "No Response," and (SDs), in the Four Fragment Conditions.

| | Syllable condition | Pronounceable nonsyllable | Unpronoun. cluster | Nonadjacent letters |
|---|---|---|---|---|
| Reaction time | 11.6 | 16.4 | 19.0 | 20.9 |
| | ( 4.2) | ( 4.7) | ( 4.1) | ( 3.6) |
| Percent of | 24.7 | 40.1 | 50.9 | 54.9 |
| no response | (15.2) | (18.4) | (15.5) | (16.0) |

Planned comparisons were performed only between those groups of words in each condition for which the fragment clusters were at comparable positions within the words. Thus, the results were based strictly on the effect of the fragment type without being confounded with position effects. The results of the planned comparisons are presented in Table 2.

## Table 2

### Planned Comparisons of Reaction Times in Seconds, between Pairs of Fragment Conditions, with Subject (SR) and Word (WR) Random.

| Conditions compared | Mean percent of no response | Mean RT (SD) | | t value |
|---|---|---|---|---|
| Syllable - | 22.4 | 11.2 | SR | $t(63) = 5.16p < 0.001$ |
| vs. | (16.4) | ( 4.5) | | |
| Pron. Nonsyl.- | 37.5 | 15.5 | WR | $t(35) = 2.53p < 0.02$ |
| | (20.7) | ( 5.2) | | |
| Pron. Nonsyl.- | 48.4 | 18.6 | SR | $t(63) = 0.89p <$ n.s. |
| vs. | (21.0) | ( 5.2) | | |
| Unpron. Clus.- | 44.8 | 17.9 | WR | $t(35) = 0.80p <$ n.s. |
| | (17.3) | ( 4.7) | | |
| Pron. Nonsyl.- | 37.1 | 15.5 | SR | $t(63) = 8.08p < 0.001$ |
| vs. | (20.7) | ( 5.2) | | |
| Nonadj. Lett.- | 55.6 | 21.0 | WR | $t(35) = 3.80p < 0.001$ |
| | (17.7) | ( 4.1) | | |
| Unpron. Clus.- | 55.5 | 20.6 | SR | $t(63) = 2.72p < 0.008$ |
| vs. | (22.2) | ( 5.2) | | |
| Nonadj. Lett.- | 64.8 | 23.0 | WR | $t(23) = 2.38p < 0.026$ |
| | (22.5) | ( 5.0) | | |

The results clearly demonstrate that the syllabic fragments are better retrieval cues than any other fragment in a given position in the word. It is of interest to note that there was no significant difference between the two kinds of nonsyllabic clusters: the pronounceable nonsyllable and the unpronounceable cluster. However, it is clear that clustering in itself facilitates retrieval, as any cluster yielded better performance than the nonadjacent letters.

While considering the facilitation of syllabic fragments versus pronounceable nonsyllabic fragments, one cannot disregard the fact that for many words the division into syllables is controversial. Although in English some words have clear syllabic boundaries( e.g., "after"), for many words the syllabic boundaries are not well defined (e.g., "dagger"). These words contain ambisyllabic segments in most cases, in which a clear and unequivocal break does not exist. Ambisyllabicity is the major cause for having more than one theory of syllabification in English, because different parsings into syllables can be suggested for many words (see Kahn, 1976).

The issue of ambiguous syllabification is not only a linguistic issue, but also a psychological and methodological one. It might be the case that some of the controversy that revolves around the effect of syllables in word perception is due to the use of stimuli whose syllabification is ambiguous. One may suggest that the use of such stimuli might have prevented the researchers from finding a clear facilitation for syllabic units. In the present study, however, we found a strong facilitation of syllabic clusters even though a great number of the experimental stimuli contained ambisyllabic segments. We believe that even greater facilitation can be demonstrated while using only words that have unequivocal syllabic boundaries.

Unambiguous syllabifications can be easily differentiated from ambiguous ones. Although linguists disagree about the correct syllabic boundaries of many words, there is a set of syllabification rules that they do agree upon. For exemple, it is fairly accepted that a syllable must begin and end with consonants or sequences of consonants that are legal in word-initial and word-final position, or that adjacent vowels belong to differnt syllables, or that the stressed syllable will contain the maximal permissable number of consonants.

Given the great theoretical relevence of syllabification ambiguity, we examined the results separately for those words whose syllabification is unambiguous. The differences between the syllabic and the nonsyllabic pronounceable clusters only increased: RT=9.6 (SD=5.5), and RT=14.9 (Sd=9.2); for syllabic and nonsyllabic fragments, respectively. The results of percentage of "no response" were similar: 16.2% for syllables, and 36% for nonsyllabic fragments.

## EXPERIMENT 2

The results of Experiment 1 showed that syllables facilitate retrieval of words from semantic memory. However, it is not clear whether the facilitation that was found for syllabic units should be attributed to phonology or to morphology. Experiment 2 was designed to address this issue by investigating the relative facilitative effect of phonologic units versus morphemic units.

Chomsky and Halle (1968) suggested that morphemes rather than phonologic units are stored in lexical memory in English. This suggestion is based on the claim that the syllabic structure of a word changes in a systematic way when affixes are added to it, while the underlying morphemic structure remains the same. Thus, it is more parsimonious to store the morphemic structure together with the rules for generating the phonologic structure according to the affixes added to the basic word.

Another source of evidence supporting the existence of morphemic units derives from reading research. Marcel (1980) suggested that in the process of reading, the reader parses the letter string not only by a cumulative and exhaustive procedure, but also according to morphemic specifications that are in the visual lexicon. Kay and Marcel (1981) presented subjects with nonwords containing legal morphemes and demonstrated that naming latencies depended on their pronounciation regularity. Kay and Marcel therefore suggested that morphemic units are probably the basis of generating phonology in beginning readers.

A different technique for investigating lexical units is suggested by Prizmental, Treiman, and Rho (1986). They presented subjects with a target letter followed briefly by a string of colored letters. Prizmental et al. demonstrated that subjects sometimes report seeing letters and colors in incorrect combinations (illusory conjunction). Hence, they investigated in what type of letter combinations these illusory conjunctions are more likely to occur. Their results suggested that syllables defined by purely phonological principles did not affect feature integration. Contrarily, syllables that were defined by morphological boundaries were functional units in the visual analysis.

However, morphemic and syllabic units tend to overlap to a great extent. In most English words the morphemic units are either identical with the syllabic units or else have one more letter at the end. This overlapping of units may be one of the reasons for the difficulty in obtaining clear-cut results concerning their effects. Therefore, to test this, in Experiment 2 we employed stimuli that contain morphemic and syllabic units that do not overlap.

## Methods

**Subjects**. Forty-eight undergraduate students from the Hebrew University, all native English speakers, participated in the experiment for course credit or for payment.

**Stimuli and design**. The stimuli were 24 English words: 7 nouns, 4 verbs, and 13 adjectives. Twenty-one of the words had four syllables, while the remaining three had five. The words were seven to twelve letters long. Their frequencies, according to Kuçera and Francis (1967), ranged from 0 to 43, with a median of 7. All the words were of Greek or Latin origin, and their decomposition into morphemes was defined according to Aronoff (1976). In order to avoid a confounding with the fragment position within the word, only the middle fragments were used as cues. Each word contained a middle morphemic unit and a middle syllabic unit that was not contained within the morphemic unit. Words of this type are words that are not pronounced according to their morphemic structure. For example, the morphemes of the word "monotonous" are "mono," "ton," and "ous," while the stressed syllable (which was the phonetic unit used in every case) is "not." We could therefore compare the effects of "_ _ NOT _ _ _ _ _" and "_ _ _ _ TON _ _ _" as cues, together with the semantic synonym: "boring; dull." Each cue was presented with a morphemic fragment to one group of subjects and with a syllabic fragment to another group of subjects. Altogether, the subjects in each group saw each word only once. They were presented with half of the syllabic fragments and half of the morphemic fragments, randomly selected. The procedure and apparatus were identical to those in Experiment 1. The list of target words is presented in the Appendix.

**Results and Discussion**

The mean retrieval time and the percentage of "no answer" for words cued by morphemic fragments and for words cued by phonetic fragments are presented in Table 3.

**Table 3**

**Mean Reaction Time in Seconds, Percentage of "No Response," and (SDs), for Words Cued by Morphemic and Phonetic Fragments.**

| | Morphemic fragment | Phonetic fragment |
|---|---|---|
| Reaction Time | 16.3 | 13.3 |
| | ( 3.5) | ( 4.2) |
| "No response" | 40.1% | 29.2% |
| | (14.0) | (16.5) |

The differences in reaction times were significant with subjects as random variable, and with words as random variable: $t(47) = 5.23, p < 0.001; t(23) = 1.92, p < 0.065$, respectively.

Experiment 2 thus showed that, at least for those words used in the study, syllabic units are more facilitative for the retrieval of words than are morphemic units. These results aparently conflict with findings in experiments that employed lexical decision and naming tasks and yielded better performance for words that were parsed according to morphemic principles (e.g., Murrell & Morton, 1974; Taft, 1979). This discrepency in results deserves attention.

The comparison of morphemic and syllabic units in English is methodologically problematic, as the results are heavily dependent on the choice of units in each experiment. The morphemic units that were used by Taft (1979) or Murrell and Morton (1974) consisted of independent lexical units (i.e., ordinary words of the language). Therefore, there is no question that these units are stored as such in the internal lexicon, and for those specific words it is reasonable to assume that the morphemic units convey more information that any other units.

The empirical question that we addressed in this experiment refers to the comparison of morphemic and syllabic units that do not have an independent lexical status. However, as was previously pointed out, in most of these cases the syllabic and the morphologic segmentations overlap. Hence, the only set of stimuli that allows one to test the relative facilitation of phonologic and morphemic units is the one that does not confound syllables and morphemes. Unfortunately, this set of words is usually comprised of words of Greek or Latin origin, and the naive reader is usually unaware of the morphemes' meaning. The results of Experiment 2 clearly demonstrate, at least for these type of words, that morphemic units do not play an important role. These units are theoretical constructs used by linguists to explain the structures of English words. Our results suggest that people do not have a deep linguistic knowledge of their language. Units that do not have a phenomenological reality for the individual do not have a psychological reality.

In conclusion, although our results do not rule out the possibility that some morphemes might be better cues, they conflict with a strong version of morphemic lexical structure that claims that only morphemes are stored in the lexicon. The pattern of cue facilitation obtained in Experiment 2 suggests that phonologic units do play a role in the retrieval of words and, all other things being equal, they are better cues for word retrieval. Since phonologic units have also been shown to play a role in the perception of both auditorily and visually presented words (e.g., Mehler et al., 1981; and Spoehr & Smith, 1975, respectively), they are thus seen to be involved in many aspects of the internal processing of words.

## GENERAL DISCUSSION

In the present study we investigated the nature of word units in the internal lexicon by using a crossword puzzle paradigm. Experiment 1 showed that any grouping of letters is more facilitative than dispersed letters in retrieving words from memory. This result, however, is not surprising. It appears that the information afforded by a given set of clustered letters is more than the sum of the information afforded by each of the cluster's constituents alone. This conclusion is in accordance with McClelland and Rumelhart's model of word recognition (1981). According to their model, the greater activation of three adjacent letters derives from the pattern of activation characteristic of any adjacent positions. The claim, however, for the existence of units in the lexicon does not refer only to the relative position of letters at the letter level, but also to the existence of independent subunits above the letter level but below the word level.

The controversy resides in the definition of these units. The results of Experiments 1 and 2 taken together demonstrate that phonologic units are more facilitative for the retrieval of words than are any other units. It is important to note that this effect cannot be attributed to pronounceability factors alone. In Experiment 1, there was no significant difference between the nonsyllabic pronounceable and unpronounceable clusters; moreover, the syllabic cluster facilitated retrieval more than either one of them.

In Experiment 2, we directly tested the relative facilitation caused by syllabic and morphemic units. Although we cannot rule out the possibility that morphemic units also play some role in the internal processing of words, we suggest that syllabic units are more central. Thus, we propose that syllabic units are stored as such in the lexicon.

A model based on this hypothesis can be constructed as an extension of the interactive model of the lexicon proposed by McClelland and Rumelhart (1981). Using similar principles, we too propose a model in which words are connected by excitatory links to the letters they are composed of. However, we suggest that the word and letter nodes are mediated by a third level that is comprised of letter units. These units reside between the word level and the letter level and are organized according to syllabic principles. According to this model, a word can be recognized or retrieved on the basis of the isolated letters contained in it. However, retrieval is facilitated if the intermediate syllabic units are activated by a previously presented cue. This is because the syllabic units are more closely related to the word level than are the dispersed letters. In the crossword puzzle task, when a syllabic configuration is presented to the solver, it directly activates the node in the lexical network that is consistent with the presented information. This node, however, only rarely activates a single word node, as usually more than one word contains one specific syllable. If the word cannot be retrieved, then the addition of semantic information may eliminate some of the possible word candidates and may cause greater activation in the remaining ones. The complete activation of a specific word in the network (i.e., the retrieval of that word) is aided, therefore, by the additional semantic cue. The semantic information that is given with the letter configuration activates in parallel, through top-down processes, those word nodes that are consistent with it. The combination of the unit's bottom-up activation and the semantic information's top-down activation finally enables the retrieval of the target word from the lexicon. By the same argument, the addition of any single letter to the letter configuration will also narrow the number of competing words, thus facilitating retrieval. If, however, the added letter completes a syllabic unit, the increase of bottom-up activation will be comprised of two factors: (1) the added activation of that specific letter but also, and more importantly, (2) the additional activation of the completed syllabic unit. Thus, the completion of a full syllabic unit increases the probability of word retrieval.

Note that although the stimuli in the present experiments were presented in the visual modality, by no means do we suggest that only the visual lexicon is involved in the process of word retrieval. As the retrieval task requires relatively long reaction times, and may not tap on-line processing, it is reasonable to believe that both the auditory and the visual lexicons are involved in the task. In many cases, the final activation of a word node (i.e., report word retrieval) can derive from activation of either one of the lexicons or both. Regardless of this possiblility, we believe that the differences in the relative facilitation of the visually presented letter clusters reflect their relative lexical status.

In conclusion, we suggest that the word-fragment completion task is a sensitive test for investigating lexical structure. Results from this task suggest that subunits of words that are larger than the letter unit are probably stored in the mental lexicon along with the words themselves. These subunits and their interconnections make up the lexical word. As syllables appear to be the best cue for word retrieval, we suggest that syllabic units have a strong lexical reality. The exact formal definition of the syllabic units in many English words is the source of large disagreement among linguists. This question, however, might be regarded as an empirical and psychological question. Thus, the word-fragment completion task could provide empirical evidence that might influence current linguistic theories.

## References

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical acess? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 340-357.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Dolinsky, R. (1973). Word fragments as recall cues: Role of syllables. *Journal of Experimental Psychology, 97*, 272-274.

Hansen, D., & Rodgers, T. S. (1965). An exploration of psycholinguistic unit in initial reading. *Proceedings of the Symposium on the Psycholinguistic Nature of the Reading Process*. Detroit: Wayne State University.

Horowitz, L. M., White, M. A., & Atwood, D. W. (1968). Word fragments as an aid to recall: The organization of a word. *Journal of Experimental Psychology, 76*, 219-226.

Kahn, D. (1976). *Syllable-based generalization in English phonology*. Doctoral Dissertation, University of Massachussets. (Reproduced by the Indiana University Linguistic Club.)

Kay, J., & Marcel, A. (1981). One process, not two, in reading aloud: Lexical analogies do the work of non-lexical rules. *The Quarterly Journal of Experimental Psychology, 33a*, 397-413.

Kuçera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Lima, S. D., & Pollatsek, A. (1983). The Basic Orthographic Syllabic Structure (BOSS) reconsidered: Lexical access via an orthographic code? *Journal of Verbal Learning and Verbal Behavior, 22*, 310-322.

Marcel, A. J. (1980). Surface dyslexia and beginning reading: A revised hypothesis of the pronounciation of print and its impairements. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds), *Deep dyslexia*. London: Routledge and Kegan Paul.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375-407.

Mehler J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20*, 298-305.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology, 102*, 963-968.

Prizmental, W., Treiman, R., & Rho, S. H. (1986). How to see a reading unit. *Journal of Memory and Language, 25*, 461-475.

Segui, J. (1984). The syllable: A basic perceptual unit in speech processing? In H. Bouma & D. G Bouwhis (Eds.), *Attention and performance X: Control of language processes*. Hillsdale, NJ: Lawrence Erlbaum.

Spoehr, K. T. (1978). Phonological recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 17*, 127-141.

Spoehr, K. T., & Smith, E. E. (1973). The role of syllables in perceptual processing. *Cognitive Psychology, 5*, 71-89.

Spoehr K. T., & Smith, E. E. (1975). The role of orthographic and phonotactic rules in perceiving letter patterns. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 21-34.

Taft, M. (1979). Lexical access via an orthographic code: The Basic Orthographic Syllabic Structure (BOSS). *Journal of Verbal Learning and Verbal Behavior, 18*, 21-39.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14*, 638-647.

Underwood, B. J., & Schulz, R. W. (1960). *Meaningfulness and verbal learning*. Chicago: J. B. Lippincott.

*Webster's New World Dictionary of the American Language*. (1964). New York: World.

## APPENDIX

### Stimuli used in Experiment 1

| Synonym | Syllable | Pronounc. nonsyllab. | Unpron. cluster | Nonadj. letters | Target |
|---|---|---|---|---|---|
| liquid metal | MER____ | _ERC___ | __RCU__ | M_R_U__ | MERCURY |
| uninhabited place | WIL_______ | _ILD______ | __LDE_____ | W___E_N___ | WILDERNESS |
| unpaid worker | VOL______ | __LUN____ | ____NTE__ | V__U_T___ | VOLUNTEER |
| pierce | PEN______ | __NET____ | ___ETR___ | P_N___A__ | PENETRATE |
| enchant | CAP______ | ___TIV___ | __PTI____ | C__T__A__ | CAPTIVATE |
| roast | BAR_____ | ___BEC__ | __RBE___ | B_R_E___ | BARBECUE |
| careless | NEG______ | ___LIG___ | _EGL_____ | N__L__E__ | NEGLIGENT |
| move around | CIR______ | ___CUL___ | __RCU____ | C_R___A__ | CIRCULATE |
| invent | FAB______ | ___RIC___ | _ABR_____ | F__R__A__ | FABRICATE |
| agreement | HAR____ | ___MON_ | __RMO__ | H_R_O__ | HARMONY |
| true | FAC____ | _ACT___ | __CTU__ | F_C_U__ | FACTUAL |
| enlarge | MAG____ | ___NIF_ | _SGN___ | M_G_I__ | MAGNIFY |
| copy | __PLI____ | DUP______ | _UPL_____ | __P_I__T_ | DUPLICATE |
| disgust | __VUL____ | REV______ | ____LSI__ | __V_L_I__ | REVULSION |
| aspect | __MEN____ | DIM______ | ____NSI__ | __M_N_I__ | DIMENSION |
| protective | __FEN____ | DEF______ | ____NSI__ | __F_N_I__ | DEFENSIVE |
| unwilling | __LUC____ | REL______ | ____CTA__ | __L_C_A__ | RELUCTANT |
| unbiased | __PAR____ | IMP______ | _MPA_____ | __P_R_I__ | IMPARTIAL |
| leavetaking | __PAR____ | DEP______ | ____RTU__ | __P_R_U__ | DEPARTURE |
| amusement | __VER____ | DIV______ | ____RSI__ | __V_R_I__ | DIVERSION |
| loathsome | __PUL____ | REP______ | ____LSI__ | __P_L_I__ | REPULSIVE |
| resentful | __DIG____ | IND______ | _NDI_____ | __D_G_A__ | INDIGNANT |
| choosy | __LEC____ | SEL______ | ____CTI__ | __L_C_I__ | SELECTIVE |
| lonely state | __CLU____ | SEC______ | _ECL_____ | _E_L_S___ | SECLUSION |
| irresistible force | ___PUL____ | _OMP______ | __MPU_____ | _O_P__S___ | COMPULSION |
| continual | ___SIS____ | _ERS______ | __RSI_____ | __R_I_T___ | PERSISTENT |
| confused | __WIL_____ | ___ILD____ | ____LDE___ | __W_L_E___ | BEWILDERED |
| forecast | ___DIC____ | _RED______ | _____CTI__ | ___D_C_I__ | PREDICTION |
| thorough | __TEN____ | ___ENS___ | ____NSI__ | __T_N_I__ | INTENSIVE |
| crowded condition | ___GES____ | _ONG______ | __NGE_____ | _O_G_S____ | CONGESTION |
| not wanted | __WEL____ | ___ELC___ | ____LCO__ | __W_L_O__ | UNWELCOME |
| spiteful | ___DIV____ | ____ICT___ | __NDI_____ | __N_I_T___ | VINDICTIVE |
| friend | ___PAN___ | _OMP_____ | __MPA____ | _O_P_N___ | COMPANION |
| repay | ___PEN____ | ____ENS___ | __MPE_____ | _O_P_N____ | COMPENSATE |
| stamina | __DUR____ | ____RAN__ | _NDU_____ | __D__A_C_ | ENDURANCE |
| excellence | ___FEC____ | ____ECT___ | __RFE_____ | ___F_C_I__ | PERFECTION |
| hurdle | _____CLE | ___TAC__ | ____ACL_ | _B___C_E | OBSTACLE |
| increase | _____PLY | _ULT____ | ____IPL_ | __L__P_Y | MULTIPLY |
| retaliation | _____SAL | ___RIS__ | _EPR____ | __P_I__L | REPRISAL |
| upright | _____CAL | _ERT____ | __RTI___ | _E_T___L | VERTICAL |
| unique | _____LAR | ___GUL__ | __NGU___ | _I_G___R | SINGULAR |
| international negotiator | _____MAT | ___LOM__ | _IPL____ | __P_O__T | DIPLOMAT |
| reference book | ____NAC | __MAN__ | _LMA___ | _L_A__C | ALMANAC |
| watchman | _____NEL | ___TIN__ | __NTI___ | __N_I__L | SENTINEL |
| buttoned sweater | _____GAN | ___DIG__ | __RDI___ | _A_D___N | CARDIGAN |
| cruel | ____MAN | __HUM__ | _NHU___ | _N_U__N | INHUMAN |
| biased | _____SAN | ___TIS__ | __RTI___ | __R_I__N | PARTISAN |
| deviant | _____MAL | ___ORM__ | ____RMA_ | ___O_M_L | ABNORMAL |

## Stimuli used in Experiment 2

| Synonym | Phonetic unit | Morphemic unit | target |
|---|---|---|---|
| not pertinent | --REL----- | ----LEV--- | IRRELEVENT |
| disrespectful | --REV----- | ----REV--- | IRREVERENT |
| manage skillfully; control | --NIP----- | ----PUL--- | MANIPULATE |
| boring; dull | --NOT----- | ----TON--- | MONOTONOUS |
| meat-eating | ---NIV----- | -----VOR--- | CARNIVOROUS |
| grow or spread rapidly | ---LIF----- | -----FER--- | PROLIFERATE |
| exclusive control or ownership | --NOP--- | ----POL- | MONOPOLY |
| disloyalty; unfaithfulness | ----DEL--- | --FID----- | INFIDELITY |
| component structure; dissection | -NAT--- | ---TOM- | ANATOMY |
| all-powerful | --NIP----- | ----POT--- | OMNIPOTENT |
| splendid | ---NIF----- | -----FIC--- | MAGNIFICENT |
| tightly joined | --SEP------ | ----PAR---- | INSEPARABLE |
| equipment | ----RAT-- | --PAR---- | APPARATUS |
| look forward to | --TIC----- | ----CIP--- | ANTICIPATE |
| independence | --TON--- | ----NOM- | AUTONOMY |
| kind; generous | --NEV----- | ----VOL--- | BENEVOLENT |
| hesitant; unable to decide | --RES----- | ----SOL--- | IRRESOLUTE |
| vague; not exact | --DEF----- | ----FIN--- | INDEFINITE |
| mix uniformly | --MOG----- | ----GEN--- | HOMOGENIZE |
| vigorous; full of pep | ----GET-- | --ERG---- | ENERGETIC |
| conflicting feelings | --BIV------ | ----VAL---- | AMBIVALENCE |
| secret; not to be disclosed | -----DEN---- | ---FID------ | CONFIDENTIAL |
| applied science | ----NOL--- | ------LOG- | TECHNOLOGY |
| unlawful | ----GIT----- | --LEG------- | ILLEGITIMATE |

# ON THE POSSIBLE ROLE OF AUDITORY SHORT-TERM ADAPTATION IN PERCEPTION OF THE PREVOCALIC [m]-[n] CONTRAST*

**Bruno H. Repp**

**Abstract.** *Acoustic information about the place of articulation of a prevocalic nasal consonant is distributed over two distinct signal portions, the nasal murmur and the onset of the following vowel. The spectral properties of these signal portions are perceptually important, as is their relationship (the pattern of spectral change). A series of experiments was conducted to investigate to what extent relational place of articulation information derives from a peripheral auditory interaction, viz., short-term adaptation caused by the murmur. Experimental manipulations intended to disrupt the effects of such adaptation included separation of the murmur and the vowel by intervals of silence, presentation to different ears, and reversal of order. Other tests of the possible role of adaptation included manipulation of murmur duration, murmur-vowel cross-splicing, and high-pass filtering of the excised vowel onset. While the results of several experiments were compatible with the peripheral adaptation hypothesis, others did not support it. An alternative hypothesis, that the manner cues provided by the murmur are crucial for accurate place judgments, was also discredited. It was concluded that, at least under good listening conditions, the perception of spectral relationships does not depend on peripheral auditory enhancement and probably rests on a central comparison process.*

## INTRODUCTION

The present study continues recent research on the perceptual integration of nasal murmur and vowel onset cues to the [m]-[n] distinction in CV syllables (Kurowski & Blumstein, 1984; Repp, 1986). Kurowski and Blumstein showed that each of these signal portions may carry considerable

---

* *Journal of the Acoustical Society of America*, in press.

*Acknowledgment.* This research was supported by NICHD Grant HD-01994 and by BRS Grant RR-05596 to Haskins Laboratories. I am grateful to Katyanee Svastikula and Hwei-Bing Lin for extensive help with stimulus preparation and running subjects. Critical comments by Quentin Summerfield and an anonymous reviewer contributed substantially to the improvement of the manuscript. Results from Experiment 4 were reported at the Twenty-Sixth Annual Meeting of the Psychonomic Society in Boston, MA, in November 1985.

place of articulation information, and that subjects' identification performance is better when both are present (as they normally are) than when only one is. They suggested that the two cues may function as a "single auditory property." However, their data also seemed consistent with the alternative possibility that the two cues are processed separately and combined at a later, evaluative stage in perception (see, e.g., Massaro & Oden, 1980). Repp referred to these two hypotheses as single-cue (or early integration) and multiple-cue (or late integration), respectively.

In addition to replicating and extending Kurowski and Blumstein's findings using a multi-talker stimulus set, Repp made a preliminary attempt to address these two integration hypotheses. He formulated a simple probabilistic model of late information integration that predicted identification accuracy when two cues are available from identification performance for each cue presented in isolation. The predictions of the model generally fell short of the obtained identification scores, which was taken to mean that perceptual integration did occur at a relatively early stage, as hypothesized by Kurowski and Blumstein. However, the model may well have been too simple to represent the processes of cognitive information integration. Another relevant piece of information obtained in Repp's study was that murmur and vowel onset cues still appeared to be integrated better than predicted by the model (or, in any case, permitted surprisingly high identification scores) when as much as 60 ms of the waveform surrounding the point of articulatory release was replaced with noise. This finding casts doubt on the role of a peripheral integration mechanism, since such a mechanism presumably should have been more sensitive to disruption of physical continuity. However, the noise may have enabled listeners to "restore" the missing acoustic information (cf. Warren, 1984). Clearly, Repp's data were not sufficient to decide between the early and late integration hypotheses, and further research was called for.

The concept of late integration needs little justification, since separate sources of information can always be combined in cognitive decision making as long as they are available at the same time (see, e.g., Massaro & Oden, 1980). The concept of early integration is more controversial, however. According to Kurowski and Blumstein's hypothesis, murmur and vowel onset "are not represented as separate cues, but are integrated *by the auditory system* into one unitary representation" (p. 389, emphasis added). As support for this claim, they cite the physiological studies of Delgutte (1980; Delgutte & Kiang, 1984), who found in cats that the neural response to a vowel onset was altered by a preceding nasal murmur, due to short-term adaptation of auditory nerve fibers. Kurowski and Blumstein conclude from this finding that "the auditory system does not treat transitions [i.e., the vowel onset] separately from the murmur" (p. 389). However, while Delgutte's results suggest that the auditory representation of the vowel onset is not *independent* of the preceding murmur, it does not follow that the two signal components, therefore, form an auditory unit. That is, one must distinguish between early *integration*, which yields a single auditory property, and early *interaction* among stimulus portions, which may modify their auditory representations while preserving them as separate sources of information that could be integrated by a later, cognitive process. Auditory adaptation would seem to be a likely source of early stimulus component interaction, but it is not clear how it ever could merge two signal portions of very different spectral structure and considerable temporal extent. Indeed, adaptation serves to enhance spectral changes in the signal (Summerfield, Haggard, Foster, & Gray, 1984) and thus is a mechanism of *differentiation*, not of integration. Thus, early integration of the kind envisioned by Kurowski and Blumstein seems unlikely as a general auditory function. Rather, the concept seems to reflect the axiomatic belief that single auditory properties underlie phonetic distinctions. This assumption is intended to relieve the listener's perceptual system from a computational burden, which instead falls upon the investigator trying to define the critical properties (Repp, 1987b).

Instead of the early integration hypothesis, therefore, the present series of experiments is concerned mainly with the perceptual consequences of early auditory interaction—henceforth, the (auditory short-term) *adaptation hypothesis*. Auditory short-term adaptation has been amply demonstrated not only in animals' auditory nerves (see also, e.g., Abbas & Gorga, 1981; Eggermont, 1985; Harris & Dallos, 1979; Smith, 1979) but also behaviorally in humans in the form of forward masking, decay of sensation, and auditory aftereffects (e.g., Plomp, 1964; Viemeister & Bacon, 1982; Widin & Viemeister, 1979; Wilson, 1970; Zwislocki, Pirodda, & Rubin, 1959), including tasks involving phonetic judgments (Summerfield et al., 1984; Summerfield & Assmann, 1987), even though adaptation may not be the only factor contributing to these phenomena. For all we know, then, auditory adaptation occurs continuously as we listen to speech. The question is: Does it help speech perception? Summerfield et al. (1984) and Summerfield and Assmann (1987) have argued that adaptation serves to enhance regions of spectral change, and that this may increase the intelligibility of speech, especially in noisy environments. In the specific case that concerns us here, viz., prevocalic nasal consonants, significant spectral change occurs at the point of release, where the nasal murmur changes into the vowel (and also beyond that point, during the formant transitions in the vowel). The murmur thus presumably has an adapting effect on the vowel onset that is proportional to the murmur spectrum, resulting mainly in attenuation of frequencies below 1000 Hz, where the murmur has most of its energy. Since distinctive place of articulation information is located at higher frequencies, some enhancement of vowel onset cues may result from the suppression of irrelevant spectral components (cf. Danaher & Pickett, 1975; Hannley & Dorman, 1983). The transitions of the second and third formants following vowel onset may also be enhanced somewhat by the (weak) presence of these formants in the murmur. More generally, the negative aftereffect of the murmur results in a direct auditory representation of the differences in spectral amplitude between the murmur and the onset of the vowel. This direct spectral difference information may be perceptually valuable, especially for the labile [mi]-[ni] distinction (Repp, 1986).

It could be that such relational spectral information is *the* critical cue for place of articulation distinctions. (See Lahiri, Gewirth, & Blumstein, 1984.) This need not be so, however, for the murmur, as well as the later portions of the vowel, provide additional spectral (and temporal) information that may feed into a central integration process. Repp's (1986) preliminary acoustic analyses suggest that spectral difference information alone is not sufficient to distinguish [m] and [n] across all vowel contexts, at least not in an invariant fashion. It also seems to vary in perceptual importance depending on the vowel, being more essential in [-i] than in [-*a*] context, for example. Thus it may be only one of several ingredients that enter into phonetic decisions. This means that the inputs to the central decision process probably include the murmur spectrum, the spectral relationship between the murmur and the vowel onset, and the continuing pattern of spectral change during the vowel.

The present series of experiments was designed to test the adaptation hypothesis in a variety of ways. To repeat, that hypothesis states that adaptation by the nasal murmur modifies the internal representation of the vowel onset spectrum and thus makes spectral difference information directly available to the auditory system, which is important for the correct perception of place of articulation. Therefore, identification scores should drop if the effect of adaptation is reduced or eliminated. It was assumed that auditory adaptation, being a peripheral process, would be sensitive to disruptions of the physical continuity of murmur and vowel, so Experiments 1-3 introduced manipulations such as order reversal, spatial separation, and temporal separation of murmur and vowel components. If such disruptions reduced identification performance substantially, a role of peripheral adaptation in providing place of articulation information would be suggested. If they had no effect, the auditory adaptation hypothesis could be rejected. A potential problem with this approach is that it is quite possible that spectral difference information, if it is not available as the direct consequence of peripheral auditory adaptation (and even if it is), is computed at a higher level in the perceptual system with the help of auditory memory (see Summerfield & Assmann, 1987; Summerfield et al., 1984), as suggested, for example, by research on auditory profile analysis (see Green, 1983). Such a central comparison process may also be sensitive to disruptions of physical continuity, and unless such disruptions turn out to be ineffective, the outcome of the experiments will be consistent with both peripheral and central explanations. To distinguish further between these accounts, Experiments 4, 6, and 7 examined several predictions thought to be specific to peripheral adaptation, concerning the effects on intelligibility of murmur duration, murmur/vowel mismatches, and simulated spectral enhancement at vowel onset. Experiment 5 addressed two alternative hypotheses, which will be introduced at that point.

## I. GENERAL METHODS

### A. *Subjects*

Three different groups of 12 or 13 student volunteers served as paid subjects, each in a single session including several experiments. All subjects were native speakers of American English and considered themselves to be free of hearing problems.

### B. *Stimuli*

The same basic stimulus set as in Repp (1986) was used, and the earlier article may be consulted for details. Briefly, the stimuli were [m*a*, mi, mu, n*a*, ni, nu] produced by three male and three female talkers, 36 syllables in all. The syllables were low-pass filtered at 4.9 kHz, digitized at 10 kHz, and modified as required. The onsets of three pitch periods (or pairs of pitch periods, in female tokens) preceding and following the point of release were marked to serve as cutpoints in waveform editing. The temporal distance between these markers was approximately 10 ms.

### C. *Procedure*

The subjects listened in a quiet room over TDH-39 earphones at a comfortable intensity. Unless mentioned otherwise, all stimulus presentations were binaural with interstimulus intervals of 3 s. The subjects in Experiments 1-4 made a forced choice between /m/ and /n/ for each stimulus, guessing when no nasal consonant was perceived. The subjects in Experiments 5-7 used a free response set including /m, n, b, d/ and /-/ (no consonant) as explicit choices. The first group of subjects participated in Experiments 1 and 4; the second group in Experiments 3 and 2; and the third group in Experiments 5, 6, and 7, in fixed order. (The experiments were renumbered for expository reasons in this article.)

### D. *Data Analysis*

Analyses of variance were performed on overall identification scores both across subjects (averaged over talkers) and across talkers (averaged over subjects). Therefore, two F values will be reported for each effect tested.[1] Differences among individual syllables will be discussed in a qualitative fashion.

## II. EXPERIMENT 1

Experiment 1 tested the auditory adaptation hypothesis in a drastic fashion by reversing the order of the murmur and vowel components. Clearly, this manipulation eliminates any adapting effect the murmur might have on the vowel onset. Therefore, if adaptation enhances place of articulation cues, performance in the reversed condition should be much worse than when the murmur immediately precedes the vowel. On the other hand, if most of the place of articulation information results from processing the two sources of information separately and coding them in a more permanent form before central integration (e.g., as vectors of likelihoods of category membership; see Chistovich, 1985; Massaro & Oden, 1980), then their order might be less important. However, if important spectral relationships are extracted centrally, that process may well be sensitive to order also. Thus it was perhaps unlikely that no decline in performance would result from an order reversal; nevertheless, the fact that this result would provide conclusive evidence against the auditory adaptation hypothesis justified the experiment.

### A. *Methods*

The experiment included five conditions, each represented by a test sequence consisting of one randomization of the 36 stimuli. The first sequence contained the original, unaltered syllables and served as warm-up. The second sequence contained the same syllables, but with about 60 ms of the waveform surrounding the point of release excised. In other words, approximately the last 30 ms of the murmur and the first 30 ms of the vowel (each corresponding to three male or six female pitch pulses) were removed and the two truncated stimulus components were joined together. This excision was done to increase the number of errors and thus to reduce ceiling effects. The relatively abrupt change from the murmur to the vowel was thought to enhance the effect of adaptation on the remaining place of articulation cues in the vowel, or at least not to decrease it. That the truncated vowel portions, as well as the truncated murmurs, still contained considerable place of articulation information was clear from earlier data (Repp, 1986). To confirm this, and to illustrate to the subjects the nature of the separate stimulus components, the third and fourth test sequences contained the truncated murmurs and vowels, respectively, in isolation. The critical fifth sequence contained the truncated vowels *followed* by the truncated murmurs after a 300 ms silent interval. This interval was inserted to prevent the perception of postvocalic nasal consonants.

---

[1] Because of frequent perfect scores, an arcsine transformation of proportions was not used. It is believed that none of the conclusions would have changed, had such a transformation been applied.

### B. *Results and Discussion*

The results, averaged over subjects and talkers, are summarized in Table 1. Performance for the unaltered syllables was 95% correct; nearly all errors occurred with [ni]. Excision of 60 ms surrounding the release caused a 10% drop in the average score, although identification of [m*a*] and [n*a*] remained unaffected. Scores for isolated truncated murmurs and vowels were 56 and 61% correct, respectively. From these scores, Repp's (1986) simple late integration formula predicts an overall performance of 66% correct for murmurs and vowels combined, without any relational information added. Clearly, however, such relational information played a role when murmur and vowel were concatenated (condition 2): Scores were much higher than predicted. In condition 5, on the other hand, where the murmur *followed* the vowel, performance was 67% correct. This is close to the predictions of the model, and while it is marginally better than identification of isolated vowels, $F(1,11) = 4.25, p = .0636; F(1,5) = 9.50, p = .0274$, it is substantially lower than the 85% correct obtained in the second condition, $F(1,11) = 49.35, p < .0001; F(1,5) = 48.75, p = .0009$. As Table 1 shows, this latter difference was obtained for all individual syllables, even though they differed markedly in their vulnerability to truncation.

**Table 1**

**Percent Correct Scores for Individual Syllables in the Five Conditions of Experiment 1. (M = murmur, V = vowel.)**

| Conditions | Syllables | | | | | | |
|---|---|---|---|---|---|---|---|
| | [mi] | [ni] | [m*a*] | [n*a*] | [mu] | [nu] | Average |
| Full syllable | 97 | 74 | 100 | 99 | 100 | 100 | 95 |
| M + V | 68 | 64 | 99 | 99 | 89 | 92 | 85 |
| M | 56 | 47 | 65 | 49 | 61 | 58 | 56 |
| V | 51 | 49 | 58 | 71 | 57 | 81 | 61 |
| V + (300 ms) + M | 57 | 47 | 82 | 76 | 68 | 71 | 67 |

These results confirm the important perceptual role of spectral difference information. When this information is directly available, speech intelligibility is much higher than when listeners can rely only on the cognitive integration of independent sources of information. Models of speech perception that assume the integration of independent cues (e.g., Massaro & Oden, 1980) are incomplete in this respect. The results are thus consistent with the adaptation hypothesis, but they cannot be taken as direct support for it. Relational information could also be derived by a nonperipheral spectral comparison process sensitive to temporal order and/or temporal separation.

## III. EXPERIMENT 2

Before turning to finer parametric stimulus variations, the results of a second gross manipulation will be reported. The rationale for Experiment 2 was that, if adaptation takes place in the peripheral auditory system, it should be sufficient to present the stimulus components to different ears to eliminate it. Summerfield et al. (1984) found that an auditory aftereffect believed to rest on adaptation disappeared when the adapting and test stimuli were presented to opposite ears. However, any central processes that extract spectral relationships might operate on inputs from different ears. As in Experiment 1, it was perhaps unlikely that the segregation of murmur and vowel would have no effect at all on intelligibility, but the strong implications such an outcome would have for the adaptation hypothesis made the experiment worthwhile.

### A. *Methods*

The same truncated murmurs and vowels as in Experiment 1 were used. There were three conditions, each consisting of one presentation of the 36 stimuli. In contrast to Experiment 1, however, the three conditions were randomized together. Two conditions were identical with conditions 2 (truncated murmur immediately followed by truncated vowel) and 4 (isolated truncated vowels) of Experiment 1, except that presentation was monaural. In the third, "split" condition, the truncated murmur occurred on the opposite channel, immediately preceding the truncated vowel, which was on the same channel as the other stimuli. Half the subjects received the vowel portions in the left ear, and half in the right ear. No ear differences were apparent, so the data were pooled over this variable.

### B. *Results and Discussion*

Performance for the monaural murmur-vowel stimuli was 86% correct, which is similar to the score obtained (with different subjects) in Experiment 1. Performance for isolated vowels (67% correct) was somewhat higher than in Experiment 1, but matches the score obtained by Repp (1986). Performance in the novel split condition was 78% correct, significantly higher than for isolated vowels, $F(1,11) = 17.47, p = .0015; F(1,5) = 9.08, p = .0297$, but lower than for monaural murmur-vowel stimuli, $F(1,11) = 23.93, p = .0005; F(1,5) = 8.07, p = .0362$.

Differences among individual syllables may be examined in Table 2. It appears that [m-] syllables gained more from the addition of a contralateral murmur to the isolated vowel than did [n-] syllables. This is surprising in the case of [mi], whose murmur by itself conveyed very little reliable information, whereas the murmurs of [m*a*] and [mu] yielded the highest scores in isolation (see Table 1; also, Repp, 1986) and therefore were expected to make a large contribution. In the case of [n*a*] and [nu], the negligible gain may have been due to the fact that the isolated vowels were identified almost as well as the monaural murmur-vowel stimuli. The possibility of response biases cannot be ruled out.[2] If the task is considered as one of [m]-[n] discrimination within each vocalic context (e.g., if percent correct scores are computed for [m]-[n] pairs), all inconsistencies disappear, and performance in the split condition is intermediate between the other two conditions in all three vocalic contexts.

---

[2] Although it seemed at times as if isolated murmurs elicited a response bias in favor of /m/ (cf. also Malécot, 1956), this tendency may indicate that labial place of articulation is more effectively conveyed by the murmur spectrum than is alveolar place. It also depends on the original vocalic context in a way that can be rationalized by reference to speech production (Repp, 1986). It is not clear, therefore, whether a meaningful distinction between discriminability and response bias can be made.

**Table 2**

**Percent Correct Scores for Individual Syllables in the Three Conditions of Experiment 2. (M = murmur, V = vowel, / = split between ears.)**

| Conditions | Syllables | | | | | | |
|---|---|---|---|---|---|---|---|
| | [mi] | [ni] | [mɑ] | [nɑ] | [mu] | [nu] | Average |
| M + V | 72 | 75 | 100 | 83 | 92 | 93 | 86 |
| V | 49 | 53 | 79 | 81 | 57 | 86 | 67 |
| M / V | 76 | 43 | 97 | 76 | 83 | 90 | 78 |

The results suggest, then, that channel separation of murmur and vowel disrupts the extraction of spectral difference information. This is consistent with the adaptation hypothesis, but it could also be that there is a central process of spectral comparison that is sensitive to spatial separation of sound sources. The scores in the split condition seem fairly close to what one should expect on the basis of late integration of independent sources of information, so the central process responsible for that integration presumably was not affected. While the results of Experiment 2, like those of Experiment 1, do not permit rejection of any specific hypothesis, they do suggest that spatio-temporal contiguity of signal components is required for the effective detection of relational spectral cues.

## IV. EXPERIMENT 3

The obvious next step was to determine how close in time the two signal components must be for listeners to reap the benefits of spectral difference information. One of the more striking findings of Repp (1986) was that substitution of signal-correlated noise (SCN) for the 60 ms of waveform surrounding the consonantal release resulted only in a relatively small decrement in overall identification performance; the syllables [mi] and [ni] supplied virtually all the errors. Repp concluded that murmur and residual vowel onset cues were perceptually integrated across the intervening noise; that is, it appeared that spectral difference information remained largely intact.[3] This result is not necessarily damaging to the adaptation hypothesis. Short-term adaptation may last for 150 ms or more (Delgutte, 1980; Summerfield et al., 1984), and a brief broadband noise may dilute but not eliminate the effect, just as would decay of adaptation during a 60-ms silent interval. However, if this interval were extended, a substantial decrement in adaptation should be observed.

---

[3] A related result has been obtained by Whalen and Samuel (1985), who substituted a nonspeech noise for the initial 60 ms of the vowel in fricative-vowel syllables and found that classification reaction time was slowed when the fricative noise had been cross-spliced from a different vocalic context. That is, listeners detected subtle phonetic mismatches between fricative noise and vowel across a 60-ms intervening noise, just as they did when no noise was present. The detection of such mismatches may rest on the extraction of spectral difference information from the speech signal.

To test these predictions, Experiment 3 assessed identification performance for stimuli whose murmur and vowel components were separated by silent intervals of up to 240 ms duration. The use of silence rather than noise was justified by the results of another experiment, not reported in detail here, which showed that intervening signal-correlated noise, broadband noise, and silence had statistically equivalent effects.[4]

### A. *Methods*

The truncated murmur and vowel components were used again, separated by 0, 30, 60, 120, or 240 ms of silence. All five conditions were randomized together and recorded in five blocks of 36 syllables each.

### B. *Results and Discussion*

The results are summarized in Table 3. There was no decline in performance over the first 60 ms of separation. Only at the longer intervals was there a small reduction in performance. Overall, the effect of temporal separation was significant across subjects, $F(4,44) = 3.70, p = .0111$, but not across talkers. With regard to individual syllables, it can be seen that identifiability declined with silence duration for [n-] but not for [m-] syllables. This may once again have been due either to an /m/ response bias that emerged as the murmur was separated from the vowel, or it may indicate that labial place of articulation was perceptually more stable under these conditions.

**Table 3**

**Percent Correct Scores for Individual Syllables in the Five Conditions of Experiment 3.**

| Silence | Syllables | | | | | | |
|---|---|---|---|---|---|---|---|
| Duration | [mi] | [ni] | [mɑ] | [nɑ] | [mu] | [nu] | Average |
| 0 ms | 71 | 67 | 100 | 89 | 92 | 94 | 85 |
| 30 ms | 81 | 57 | 99 | 94 | 93 | 96 | 87 |
| 60 ms | 78 | 51 | 100 | 90 | 93 | 94 | 85 |
| 120 ms | 74 | 53 | 99 | 86 | 99 | 81 | 82 |
| 240 ms | 76 | 54 | 97 | 83 | 92 | 76 | 80 |

---

[4] Signal-correlated noise is spectrally uniform (Schroeder, 1968) but preserves the amplitude envelope of the replaced signal, which may aid listeners in "restoring" missing phonetic information (see Warren, 1984; Whalen & Samuel, 1985); if anything, however, the noise interfered more with consonant identification than did silence. In a recent study using similar methods, Parker and Diehl (1984) likewise found no difference between the effects of intervening noise and silence on vowel identification performance in "centerless" CVC syllables, and Whalen (1984) also found effects of fricative-vowel mismatches across an intervening 60-ms silent interval, just as he did across an intervening noise (Whalen & Samuel, 1985).

These results are not so easy to reconcile with the adaptation hypothesis. First, the decline in performance was small and did not occur with all syllables and talkers. Second, there seemed to be no decline at all over the first 60 ms of separation, although auditory adaptation, which decays exponentially (Eggermont, 1985; Harris & Dallos, 1979), should have decreased significantly in that interval. Since the truncated murmur and vowel components were in their original temporal relationship when separated by 60 ms, a perceptual advantage resulting from this fact may conceivably have counteracted any decline due to decay of adaptation at short intervals. Apparently, however, listeners still had spectral difference information available with 240 ms of temporal separation, and this suggests that they used auditory memory for the murmur to determine its spectral relationship to the vowel onset. Whether this was a compensatory perceptual strategy or whether it reflects what occurs in intact syllables is not clear.

## V. EXPERIMENT 4

Experiment 4 addressed two further predictions of the adaptation hypothesis, which contrasted with predictions arising from the alternate hypothesis that murmur and vowel onset function as independent cues that are integrated at a late stage (e.g., Massaro & Oden, 1980). One prediction concerned the effect of murmur duration. Physiological studies have shown that auditory adaptation in animals increases with adaptor duration up to about 100 ms (Harris & Dallos, 1979; Westerman & Smith, 1984). Even though the temporal parameters may not be exactly the same in the human auditory system, to the extent that auditory adaptation by the murmur enhances the spectral structure at vowel onset, there should be a beneficial effect of increasing murmur duration (up to about 100 ms) on identification of murmur-vowel stimuli. In isolated murmurs, however, there can be no such enhancing effect of adaptation; therefore, increasing murmur duration beyond some minimum should have little influence on intelligibility. This was already suggested by Repp's (1986) analysis of the effect of natural variations in murmur duration; in addition, he found that the intelligibility of truly steady-state isolated murmurs *decreased* as their duration was increased, perhaps because their artificial quality became more apparent as they got longer. Thus a statistical interaction of the effect of murmur duration with the factor of presence versus absence of a following vowel is predicted. A contrasting prediction emerges from the late integration of independent cues hypothesis: Whether increasing murmur duration increases or decreases the informational value of the murmur, it should do so regardless of the context in which the murmur occurs.

A second prediction examined by Experiment 4 was this: If auditory adaptation caused by the murmur improves perception of higher formants at vowel onset, then a beneficial effect of prefixing an isolated vowel portion with a murmur should be obtained regardless of whether or not the murmur derives from the same utterance. The reason is that all murmurs are spectrally rather similar below 1000 Hz, where most of their energy is concentrated. And although [m] and [n] murmurs differ in the frequencies of their higher formants, which are continuous with the formants at vowel onset, it may be argued that the spectral change at vowel onset would be enhanced even more if the murmur formants were different from those at vowel onset. The paradoxical prediction is, therefore, that addition of an inappropriate murmur to an isolated vowel may *improve* identification, relative to the isolated vowel baseline. The opposite result is predicted by the independent cues hypothesis: The introduction of a conflicting cue cannot possibly improve performance. (Late integration of murmur and vowel onset cues may occur following an early auditory interaction, in which case two opposing tendencies may cancel in the data.) To test these predictions, the experiment included both compatible and conflicting murmur-vowel combinations. Thus it was also possible to compare directly two types of conditions for nasal consonants that previously have been employed in separate studies (Kurowski & Blumstein, 1984; Malecot, 1956) or with other place of articulation contrasts (Recasens, 1983).

### A. *Methods*

The experiment included one long randomized test sequence composed of 9 x 36 = 324 stimuli, and a shorter sequence of 3 x 36 = 108 stimuli. The stimulus components were steady-state murmurs generated by reiterating a single 10-ms segment of the original murmurs, taken from the vicinity of the release (see the Static Excerpts condition of Repp, 1986) and vowel portions whose initial 10 ms (one male or two female pitch pulses) had been removed.[5] The first test sequence contained the vowel portions in isolation and immediately preceded by 1, 3, 6, or 12 segments of matched or mismatched murmur. The murmur durations thus were in the vicinity of 10, 30, 60, and 120 ms. The mismatched murmurs came from the syllable with the same vowel but a different consonant, produced by the same speaker. The second, shorter test sequence contained only isolated murmurs of 30, 60, and 120 ms duration. (The 10-ms murmurs were omitted because they were easily missed in listening.)

### B. *Results and Discussion*

The overall results are shown in Figure 1. The figure plots percent correct scores as a function of murmur duration for isolated murmurs and for murmur-vowel stimuli with matched and with mismatched components. (In the case of mismatched components, "correct" responses are defined with respect to the *vowel* portion.) The data point on the ordinate, corresponding to zero murmur duration, represents the score for isolated vowels (72% correct). The results indicate that addition of a 10-ms matched or mismatched murmur to the vowel changed identification performance little, whereas addition of a murmur 30 ms long or longer resulted in an improvement, but only if the murmur matched the vowel. Mismatched murmurs neither improved nor hindered identification. Isolated murmurs of 30 and 60 ms duration were identified at levels above chance, but 120-ms murmurs could not be reliably identified. This last finding (which may have been a consequence of the artificial steady-state nature of the murmurs; cf. Repp, 1986) contrasts with the differential effect of 120-ms matched and mismatched murmurs when they preceded a vowel.

A two-way analysis of variance of the scores for the murmur-vowel stimuli yielded a significant effect of match/mismatch, $F(1,11) = 19.01, p = .0011; F(1,4) = 13.31, p = .0218$, and a significant interaction with murmur duration, $F(3,33) = 6.34, p = .0016; F(3,12) = 4.28, p = .0285$, obviously due to the shortest murmur duration, whereas the main effect of murmur duration was not significant. A separate analysis of the isolated murmurs showed a significant effect of murmur duration, $F(2,22) = 3.98, p = .0335; F(2,8) = 8.85, p = .0094$, suggesting that the performance decrease for the longest murmurs was real.

---

[5] The artificial murmurs were used to have better control over murmur duration and amplitude contour, and slightly truncated vowels were employed to avoid ceiling effects in performance. The truncation was less than in Experiments 1-3, but for no stringent reason; as before, it was assumed that truncation would merely reduce the information available without changing basic auditory and perceptual processes.

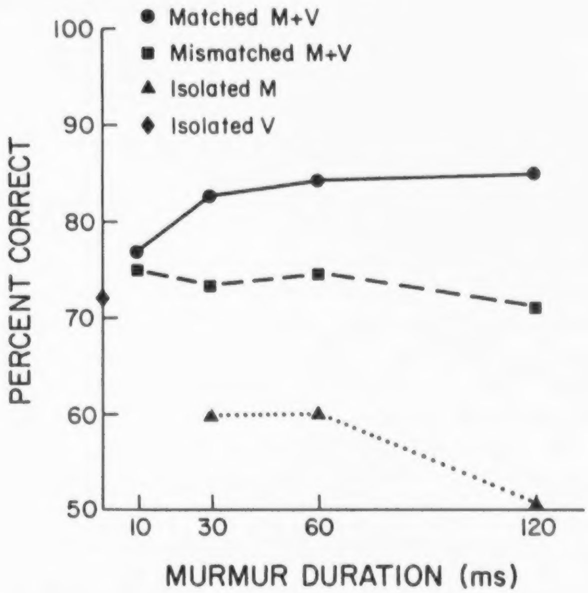

**Figure 1.** Results of Experiment 4: Percent correct identification of isolated vowels (V), isolated murmurs (M), and murmur-vowel stimuli (M + V) with matched and mismatched components as a function of murmur duration.

These overall results cannot be given much weight, however, in view of very striking dependencies on vocalic context. The pattern of results for individual syllables is shown in Figure 2. Each panel shows data for one vocalic context, with solid and open symbols representing [m-] and [n-] syllables, respectively. Consider first the [-*a*] and [-u] syllables (left and right panels). The isolated vowels of [n*a*] and [nu] were identified much more accurately than those of [m*a*] and [mu], which replicates earlier findings (Experiments 1 and 2; Repp, 1986) and probably reflects the greater perceptual salience of alveolar than labial formant transitions (or onset spectra). Because of this pattern, a [(m)*a*] or [(m)u] vowel benefited from addition of a murmur (even a 10-ms one) while a [(n)*a*] or [(n)u] vowel did not. Identification performance was uniformly high for all murmur-vowel stimuli in [-*a*] and [-u] context. Moreover, there was very little difference between scores for stimuli with matched and mismatched components. Identification of [(m)*a*] and [(m)u] vowels was *improved* by addition of a mismatched murmur almost as much as by addition of a matched murmur, and identification of [(n)*a*] and [(n)u] vowels was at least not hampered by addition of a mismatched murmur.

This part of the data is consistent with the adaptation hypothesis. As to the predicted effects of murmur duration, they are smaller than expected but are also compatible with the hypothesis. The results are inconsistent with the independent cues hypothesis, according to which performance should have decreased in the mismatched conditions.

The pattern for [mi] and [ni] stimuli (center panel of Figure 2) is very different from the results just described. Identification of isolated vowels and isolated murmurs was extremely poor, in agreement with earlier results. Addition of a 10-ms murmur to the vowel had no effect, but addition of a murmur 30 ms or more in duration elicited responses that reflected the nature of the *murmur*. Thus there was a large effect of match versus mismatch, which accounts for the average

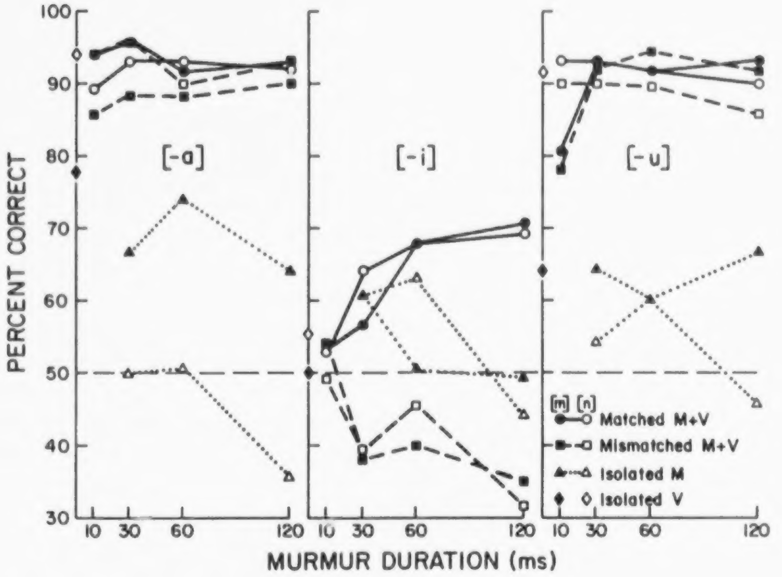

**Figure 2.** Results for individual syllables in Experiment 4.

effect shown in Figure 1. Since the murmurs were barely discriminable in isolation, especially when they were 120 ms long, listeners cannot have relied on them directly to identify the consonant in these syllables. The data thus support the earlier conclusion (Repp, 1986) that [-i] syllables are special in that place of articulation information lies almost entirely in the *relationship* between the murmur and the vowel, that is, in the pattern of spectral change. A possible implication is that there are differences between [m(i)] and [(n)i] murmurs that are difficult to detect in isolation but that become perceptually salient when the murmur is followed by a vowel. Such a retroactive enhancement effect would refute the adaptation hypothesis. Yet there is a way in which it could arise through adaptation: Different murmurs might impose their inverse spectrum on the vowel onset, thereby creating a place of articulation cue following the release. On the other hand, the independent cues hypothesis, unless it is extended to include relational information, cannot explain how murmurs that are uninformative in isolation convey phonetic differences in context.

In summary, while the results of Experiment 4 argue very clearly against the independent cues hypothesis and thereby affirm the importance of relational spectral information, they are perhaps still compatible with a peripheral account of spectral difference detection.

## VI. EXPERIMENT 5

Prior to Experiments 6 and 7, which attempted to test the adaptation hypothesis in yet another way, Experiment 5 examined two alternative explanations of how a preceding murmur might enhance the perception of vowel onset cues. One hypothesis (Repp, 1986) takes account of the fact that the murmur is the major carrier of nasal manner information. If it were the case that place of articulation perception is not independent of manner perception (see Carden, Levitt, Jusczyk, & Walley, 1981; Miller, 1977), then hearing the correct manner may enhance the accuracy of place identification. Kurowski and Blumstein (1984) reported that their CV syllables were identified as beginning with oral stop consonants when the nasal murmur was excised. Their subjects chose from the response set /m, n, b, d/ and gave about 84% /b, d/ responses to murmurless stimuli but only about 12% to stimuli with an initial murmur. Thus removal of the nasal murmur clearly changed manner of articulation perception and perhaps affected place of articulation perception as well, particularly since the isolated vowel portions of nasals lack the release bursts commonly associated with oral stop consonants. In Repp's (1986) experiments, and in Experiments 1-4, subjects always were required to make a forced choice between /m/ and /n/, regardless of whether they perceived the correct manner or indeed any consonant at all. One purpose of Experiment 5 was to determine first whether the present stimuli resembled those of Kurowski and Blumstein (1984) in that removal of the murmur resulted in the almost complete loss of nasal manner cues, and then whether correct perception of place was contingent on correct perception of manner.

A second hypothesis addressed by Experiment 5 derives from observations by Pols and Schouten (1978, 1981) on the perception of truncated stop-consonant-vowel syllables. These authors argued that the relatively abrupt stimulus onset following truncation causes spectral splatter (a "click sensation") that interferes with the perception of place of articulation cues. Identification scores improved substantially when the truncated syllables were preceded by noise bursts that masked the abrupt onset (Pols & Schouten, 1978). Ohde and Sharf (1981) applied a smoothing function to the onsets of truncated CV syllables, apparently with similar results (see Pols & Schouten, 1981). It is possible that part of the intelligibility decrement for isolated vowel portions in Experiments 1-4 was caused by abrupt stimulus onsets. To check on this, a smoothing function similar to that used by Ohde and Sharf (1981) was applied to the stimulus onsets on half the trials in this experiment.

### A. *Methods*

The experimental tape contained 8 x 36 = 288 isolated vowel stimuli in random order. Each vowel was truncated approximately 0, 10, 20, and 30 ms after the release (see Repp, 1986); thus none of them contained any nasal murmur. (It was quite clear from informal listening that inclusion of even a very brief murmur resulted in the perception of a nasal consonant.) Each truncated stimulus occurred in two versions, one unaltered and the other with a linear amplitude ramp, rising from near-zero to full intensity in 10 ms, applied to the onset of the digitized waveform. The subjects' task was to report for each stimulus the initial consonant they heard, choosing from the set /m, n, b, d/, and to write down a dash when no consonant was heard.

### B. *Results and Discussion*

The overall results, averaged over the ramped and unramped stimulus versions, are shown in Figure 3. Three measures were derived from the data. The first, p(C), was the percentage of trials on which a consonant was reported. Not surprisingly, it declined with progressive truncation, $F(3,36) = 47.05, p < .0001; F(3,15) = 94.55, p < .0001$, although the vowel portions were still heard as containing initial consonants on about half the trials even after their initial 30 ms had been deleted. The other two measures were conditional on a consonant being reported. The percentage of correct place identifications, $p_c(P|C)$, declined only very slightly with truncation, $F(3,36) = 2.17, p = .1081; F(3,15) = 6.45, p = .0051$, suggesting that the decrease in two-alternative forced-choice identification scores with progressive truncation (Repp, 1986) was caused

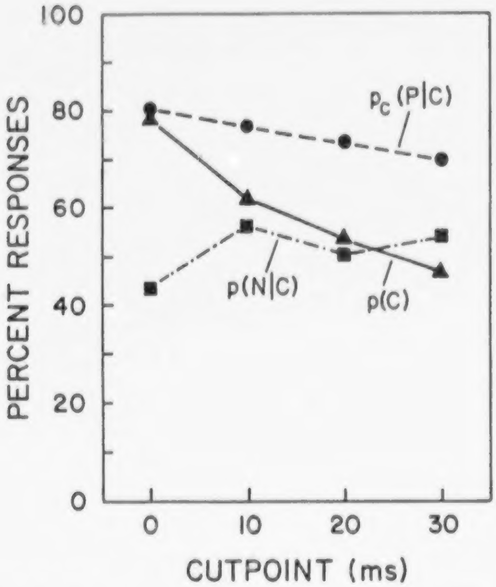

**Figure 3.** Results of Experiment 5: Percentages of consonant responses, p(C), of correct place of articulation identifications given a consonant response, $p_c(P|C)$, and of nasal consonant responses given a consonant response, p(N|C), as a function of cutpoint location.

more by the total loss of consonantal cues than by misleading residual cues. Most interestingly, the percentage of nasal consonant responses, p(N|C), did not decline at all with progressive truncation, but actually showed an initial increase, $F(3,36) = 9.72, p = .0001; F(3,15) = 10.22, p = .0006$. Regardless of how much consonantal information was available, about half of the consonants reported were nasals. This percentage is much higher than that reported by Kurowski and Blumstein (1984), even though removal of the nasal murmur undoubtedly caused a significant loss of nasal manner information. Presumably, the talker used by Kurowski and Blumstein closed his velum more rapidly after the consonantal release than did the present talkers, who tended to nasalize the vowel onset.

Differences among individual syllables are shown in Figure 4. With regard to the percentage of consonant responses (left panel), it can be seen that [mu] and [ni] were affected much more by excision of the murmur (0 ms cutpoint) than the other syllables. This probably reflects the weak formant transitions in these stimuli, which have similar articulatory configurations for consonant and vowel. Further truncation had especially strong effects on [m*a*] and [mi], indicating the loss of rapid labial transients at stimulus onset. Perception of the consonants in [n*a*] and [nu], which have relatively long vocalic formant transitions, was most resistant to vowel truncation.

The most striking difference in correct place of articulation identification scores (center panel) was between [ni] and all other syllables. Without the murmur, [ni] tended to be misidentified as labial, which indicates that the vowel did not contain any useful formant transition information. The same may well be true for [mu], and the 70-80% labial responses to both of these syllables may represent a bias to respond with labial consonants in the absence of clear place of articulation cues. Only [m*a*] and [mi] were affected by vowel truncation.

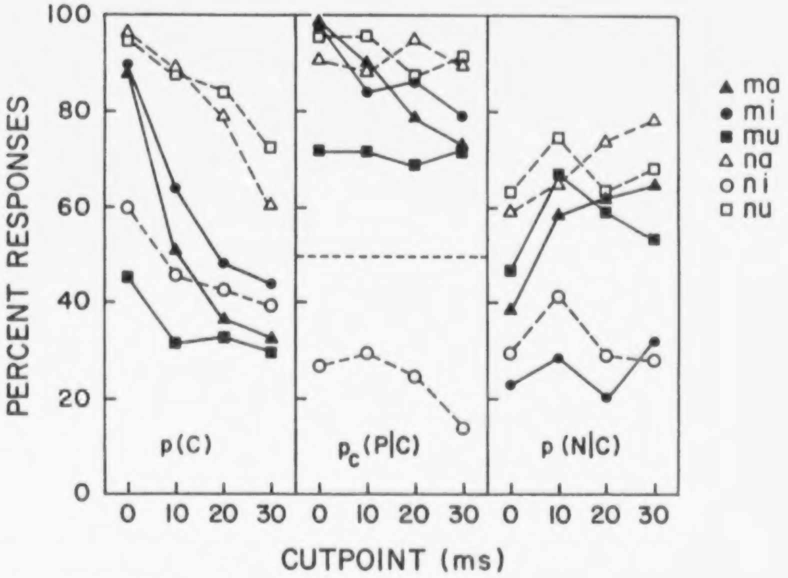

**Figure 4.** Results for individual syllables in Experiment 5.

The percentage of nasal responses (right panel) was lower for [mi] and [ni] than for the other syllables. The difference between [-i] and [-*a*] syllables may be explained by the fact that the velum is raised faster for high than for low vowels following a nasal consonant (Bell-Berti, Baer, Harris, & Niimi, 1979), making the former less nasalized. It is not clear, however, why the [-u] syllables resembled more the [-*a*] than the [-i] syllables in degree of perceived nasality, or why perception did not fully compensate for the expected differences in velar elevation for vowels of different heights (see Abramson, Nye, Henderson, & Marshall, 1981).

The principal hypothesis addressed by this experiment concerned the possible dependence of place perception on manner perception. Since only about half of the initial consonants perceived in truncated syllables were nasal, it is indeed possible that place of articulation perception suffered because of insufficient manner cues. If so, place identification *contingent* on correct perception of nasal manner should have been more accurate than place identification contingent on perception of non-nasality. Examination of these percentages (computed from the syllable averages), however, revealed only a small difference (2% on the average) in the predicted direction. This difference, moreover, derived entirely from the stimuli with tapered onsets (5.5% average difference); for the others, there was a 1.6% difference in the opposite direction. Although the effect of amplitude tapering deserves attention (see below), all stimuli in earlier experiments were, of course, untapered. For those stimuli, then, there is no evidence that incorrect perception of manner impaired place of articulation identification, so the perceptual enhancement of place cues when a vowel is prefixed with a murmur cannot be explained on that basis.

It is noteworthy, however, that there were very large differences among individual syllables. The differences between correct place identification scores contingent on perceived nasal and non-nasal manner were: -2.8% for [m*a*], -18.3% for [mi], -28% for [mu], 12.5% for [n*a*], 42% for [ni], and 6.3% for [nu]. It thus appears that, when a consonant was perceived as non-nasal, there was a strong shift in favor of labial responses; the differences in absolute magnitude of this shift among the six syllables probably derived largely from ceiling effects. Thus there *was* a dependency of place of articulation identification on manner, though in terms of criterion rather than accuracy. This effect is in agreement with earlier findings (Larkey, Wald, & Strange, 1978; Miller, 1977) of a relative shift in the category boundary on synthetic /ba/-/da/ and /ma/-/na/ continua. One likely cause for this is the absence of release bursts in both the synthetic stop-consonant-vowel stimuli used previously and in the present vowel portions. In real speech, alveolar oral stops have stronger release bursts than do labial oral stops, so the absence of bursts promotes the perception of labials, provided that a stop consonant is perceived.

Turning finally to the effect of amplitude tapering, there were small but consistent effects on two of the three overall performance measures. The percentage of consonant responses was reduced by about 7% at all stages of truncation, $F(1,12) = 16.19, p = .0017; F(1,5) = 6.12, p = .0563$, which suggests a loss of general manner cues at stimulus onset. Given that a consonant was heard, however, place of articulation identification was improved by about 5% overall, $F(1,12) = 7.60, p = .0174; F(1,5) = 11.17, p = .0205$. This effect is in agreement with the observations of Pols and Schouten (1981) on the interfering effect of abrupt stimulus onsets, although the size of the present effect was rather small—certainly much smaller than the improvement obtained by Pols and Schouten (1978) with a noise prefix. Actually, the present improvement derived solely from those trials on which nasal consonants were perceived (cf. the interaction reported above); when nasality was not perceived, there was no effect of tapering. This is less in agreement with Pols and Schouten. Onset tapering had no systematic overall effect on nasal manner perception.

In summary, the results of this experiment do not support the hypothesis that, when a vowel is preceded by its original murmur, part of the improvement in place of articulation identification derives from the restoration of correct manner identification. Perception of nasal manner does not seem to enhance perception of place, at least not in untapered stimuli as used previously; it only shifts the response criterion in favor of alveolar responses. The second hypothesis, that elimination of abrupt onsets improves place perception, receives some limited support from the present results. Though the effect is rather small, it may add to the contribution of a preceding murmur. However, it cannot explain correct perception of the intact syllable [ni], or of [mi] with truncated vowel, for which the murmur and the vowel in isolation are equally uninformative. The concept of relational information is still required, and so we must return to the adaptation hypothesis.

## VII. EXPERIMENT 6

The final two experiments in this series provided perhaps the most direct test of the adaptation hypothesis. If peripheral adaptation by the murmur enhances spectral information at vowel onset, then it should be possible to simulate this enhancement by filtering the vowel onset in the absence of a preceding murmur. Such artificial enhancement then should result in improved place of articulation identification from isolated vowel components. Confirmation of this prediction would not only provide strong support for the adaptation hypothesis, but it would also lead to a re-evaluation of earlier conclusions based on place of articulation identification from isolated vowel portions (Experiments 1, 2, and 5; Kurowski & Blumstein, 1984; Repp, 1986), which did not consider that removal of a murmur also eliminates its adaptive aftereffect.

In choosing an appropriate filtering function, decisions had to be made concerning its shape, depth, and decay over time. Acoustical analysis of the nasal murmurs indicated that most of their energy was below 1000 Hz, and that the peak corresponding to the first formant was about 30 dB higher, on the average, than the peaks of the higher formants above 1000 Hz. Only the higher formants, however, varied with place of articulation. Ideally, the spectral shape of the filter should initially mirror that of the natural murmur and then wane over time, simulating decay of auditory adaptation. These objectives were difficult to achieve simultaneously with the facilities available. In Experiment 6, therefore, it was decided to use a simple high-pass filter with a cutoff frequency of 1000 Hz, which permitted variable stop-band attenuation to simulate decay. The experiment thus tested one specific version of the adaptation hypothesis, viz., that enhancement of place cues in higher formant transitions at vowel onset results from suppression of energy in the region of the first formant. As to the decay time, it was assumed that it would be rather short during stimulation by the vowel itself. (Most estimates of decay times in the literature derive from observations during silent intervals.) Even if the range chosen (up to 30 ms) seems too short, it became clear during stimulus preparation that more extensive filtering led to very unnatural-sounding stimuli.

### A. *Methods*

The basic stimuli were the complete vowel portions of the original 36 syllables. Even though ceiling effects in performance were expected to limit the sensitivity of the experiment to beneficial (but not detrimental) effects of filtering, no truncation was performed on the vowels in this study and the next, so as to preserve the original acoustic properties of the vowel onsets. Three degrees of high-pass filtering were imposed on initial pitch-pulse segments, leaving the rest of the waveform intact: (1) the initial 10-ms segment only, with 10 dB stop-band attenuation; (2) the initial segment with 20 dB, and the following segment with 10 dB stop-band attenuation; (3) the initial segment with 30 dB, the following segment with 20 dB, and the final segment with 10 dB stop-band attenuation. Thus, three degrees of adaptation with three decay times were crudely simulated. The filtering was performed digitally, using an eighth-order elliptic filter with a fixed cut-off frequency of 1000 Hz and variable attenuation, constructed by the EFI subroutine of the ILS package (Version 4.0, Signal Technology, Inc.). The boundaries of the pitch pulses(s) to be filtered in each pass through the routine were specified precisely in tenths of milliseconds, according to Repp's (1986) cutpoint markers. The result was verified through inspection of waveforms and acoustic analysis. The four series of 36 stimuli (three filtered, one unaltered) were randomized together. Subjects were instructed to identify each stimulus as beginning with /m,n,b,d/ or /-/ (no consonant).

### B. *Results and Discussion*

The overall results are shown in Figure 5 in terms of the three performance measures introduced in Experiment 5. Looking first at the p(C) scores, it can be seen that, in agreement with the results of Experiment 5, the unaltered syllables elicited close to 80% consonant responses. This percentage declined to 65% with progressive filtering: $F(3,36) = 18.47, p < .0001; F(3,15) = 14.49, p = .0001$, suggesting that the first formant contributed general consonant manner information. A decline with respect to the unaltered stimuli was also observed in the conditional percentage of nasal consonant responses, p(N|C), $F(3,36) = 5.33, p = .0038; F(3,15) = 6.86, p = .0039$, although it did not seem to depend on the extent of filtering. Most importantly, the conditional percentage of correct place of articulation identifications, $p_c(P|C)$, also *declined*, rather than increased, with increasing extent of filtering. Although absence of an increase in performance could be blamed on ceiling effects, and although the decline is rather small and nonsignificant,

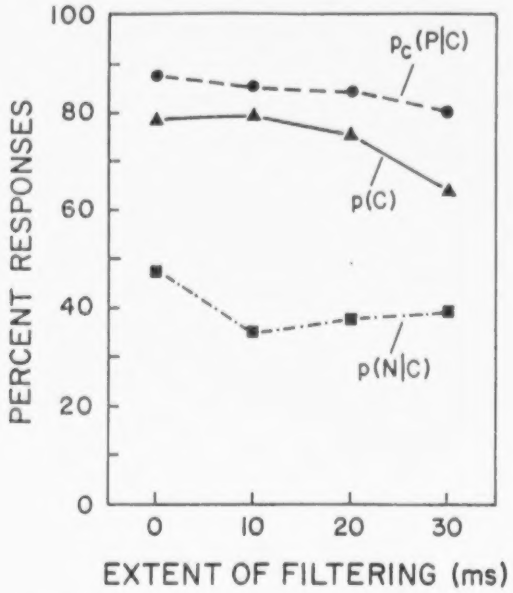

**Figure 5.** Results of Experiment 6: Three performance measures as a function of temporal extent of high-pass filtering.

these data offer no support for the hypothesis that attenuation of irrelevant low-frequency energy enhances place-of-articulation cues at higher frequencies.

Figure 6 shows the results for individual syllables. In the left panel it can be seen that consonant responses decreased most strongly for [ma] and [mi], whereas [mu] actually showed an increase with filtering. Place perception suffered in all syllables but the poorly identified [ni], for which there was an increase with filtering. Since identification of this syllable never exceeded chance level, the increase is probably a criterion effect. Perception of nasality suffered in all syllables but [ma], which showed an increase with filtering. These interactions are curious, but they do not change the general conclusions.

## VIII. EXPERIMENT 7

The results of Experiment 6 lend no support to the specific hypothesis that auditory adaptation enhances place of articulation perception through elimination of irrelevant low-frequency spectral energy. It is still possible, however, that a beneficial effect of adaptation occurs at higher frequencies, where the important place of articulation cues reside. To test this version of the adaptation hypothesis, it was necessary to use a filter that preserves the detailed spectral shape of the murmur, with some loss of flexibility in other respects.

### A. *Methods*

From each of the 36 original murmurs, a 14-coefficient LPC spectrum was computed using a 25.6 ms Hamming window ending about 10 ms before the point of release (ANA program of the ILS package). Each of these spectra was subsequently used as an inverse filter on the complete vowel portion of each syllable (FLT program). Degree of attenuation could not be varied easily in this procedure. To vary temporal extent in synchrony with pitch pulses, which could not be done

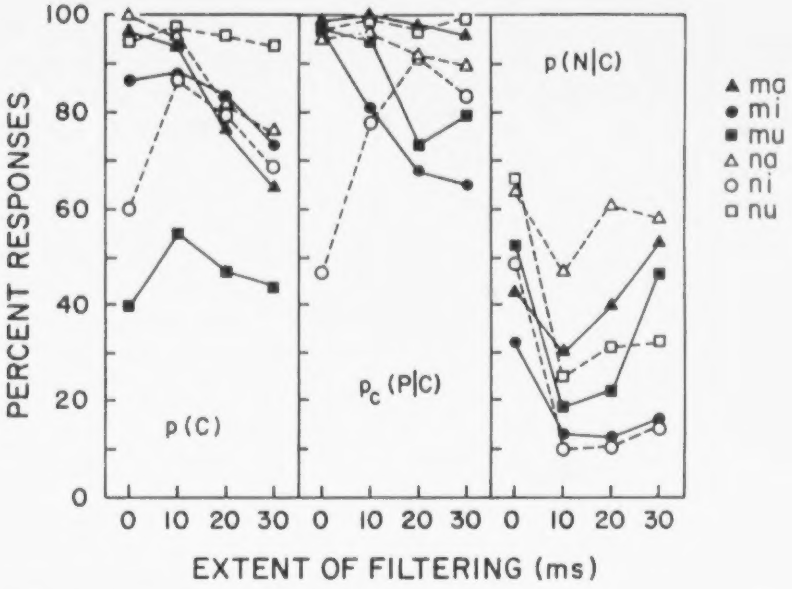

**Figure 6.** Results for individual syllables in Experiment 6.

directly, the initial one, two, or three pitch-pulse segments of the filtered vowel (about 10, 20, and 30 ms long) were concatenated with the remainder of the unfiltered vowel, using a waveform editing program. The success of the filtering procedure was verified by acoustic analysis. The resulting 4 x 36 stimuli (including the unaltered versions) were recorded in a randomized sequence. The subjects' instructions were the same as in Experiment 6. Two additional sequences of 36 stimuli each were recorded afterwards, each containing the excerpted initial 30-ms segments of the vowels, first unfiltered and then filtered. The purpose of this was to assess to what extent any perceptual effects of filtering depended on the following unfiltered vowel or were artifacts of the abrupt amplitude change between filtered and unfiltered waveform segments. In responding to these final two sequences, subjects had to make a forced choice between /m/ and /n/ for each stimulus.

### B. *Results and Discussion*

Figure 7 shows the overall results for the main test. It can be seen that the pattern was rather similar to that obtained with high-pass filtering (Figure 5). Consonant responses increased slightly initially but then decreased with increasing filtering: $F(3,36) = 13.98, p < .0001; F(3,15) = 8.91, p = .0012$. Nasal consonant responses dropped considerably with minimal filtering and then recovered partially as filtering increased: $F(3,36) = 26.79, p < .0001; F(3,15) = 16.45, p = .0001$. Correct place of articulation responses were not significantly affected, but certainly showed no tendency to increase. The results for the isolated 30-ms segments likewise showed no advantageous effects of filtering: Forced-choice identification scores were 66.5% and 64.3% for unfiltered and filtered excerpts, respectively—a nonsignificant difference.

Scores for individual syllables are shown in Figure 8. It can be seen that consonant responses increased initially for [mu] and [ni], suggesting that an initial amplitude discontinuity provides

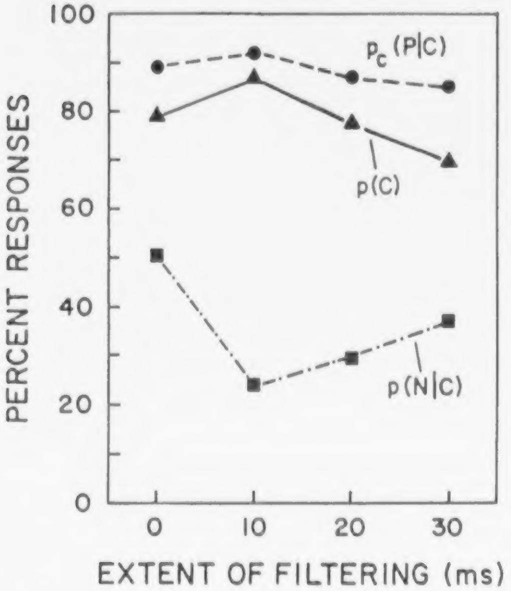

**Figure 7.** Results of Experiment 7: Three performance measures as a function of temporal extent of inverse filtering.

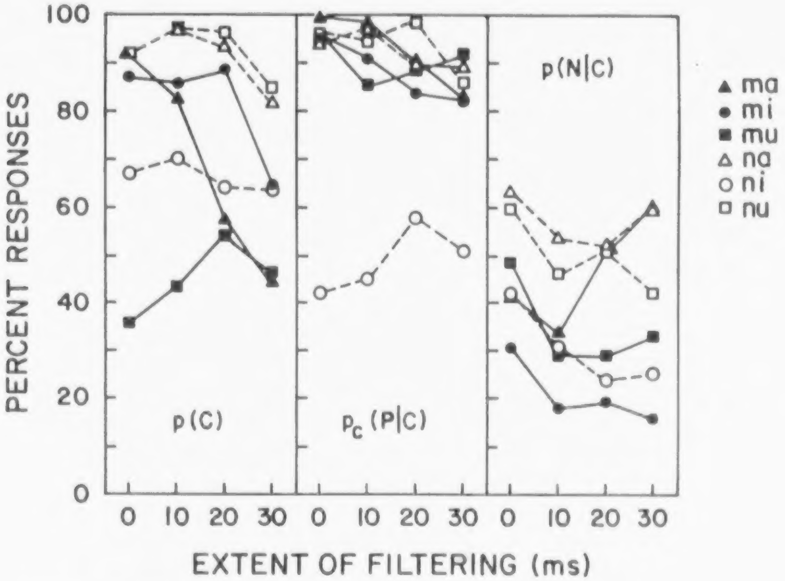

**Figure 8.** Results for individual syllables in Experiment 7.

a general consonant manner cue. With more extensive filtering, however, the cue lost its effectiveness, and consonant scores declined for all syllables. Place of articulation identification was strikingly *improved* by filtering for one syllable, [ni], but it decreased for [mi] and [mu]. The opposite effects of filtering on [mi] and [ni] suggest that, rather than improving place of articulation perception, the filtering introduced a bias to perceive /n/. No striking differences among individual syllables were observed with regard to perception of nasal manner.

In summary, these results do not support the adaptation hypothesis. It is possible, of course, that perceptual benefits of spectral enhancement are obtained only when a murmur is physically present. If so, however, the implication would be that the crucial spectral relationships are computed at a higher level, rather than being directly available in the auditory system.

## IX. SUMMARY AND CONCLUSIONS

As was already clear from earlier research, the murmur and vowel portions of nasal-consonant-vowel syllables do not make independent contributions to place of articulation perception; their relationship also plays a role. (For a recent convincing demonstration of the general importance of spectral change information in speech perception, see Furui, 1986.) This finding, which is strongly supported by the present results, argues against models of perceptual integration based on spectrographically defined cues, which do not take relational information into account. Such models have, more or less explicitly, formed the basis of much past research on speech perception (e.g., Massaro & Oden, 1980; Repp, 1982). While they may be accurate when the cues represent different (e.g., spectral vs. temporal) aspects of the speech signal, they need to be augmented by a relational term when both cues are from the same physical dimension.

The focus of the present series of experiments was the question of how listeners extract spectral relationships from the acoustic signal. That the auditory system computes some kind of running Fourier transform of the input has been an unquestioned underlying assumption. Given this assumption, there are two ways in which a listener may derive relational spectral information: directly, through auditory transforms caused by peripheral adaptation, or indirectly, through a central comparison of the spectra of successive signal portions. These two processes are not mutually exclusive: Although central comparisons seem superfluous after peripheral processes have done the work, they may substitute for peripheral processes that are artificially disrupted, and they may also serve to compute higher-order patterns of change (e.g., the second derivative of the input). The effect of adaptation in nasal-consonant-vowel syllables would be to enhance the spectral change at vowel onset and beyond. According to the strong version of the adaptation hypothesis espoused by Kurowski and Blumstein (1984), the resulting direct auditory representation of the spectral relationship would be the one and only place of articulation cue, making any further integration higher up in the system unnecessary. According to a weaker version of the hypothesis, the information obtained from the modified vowel onset is combined with cues obtained independently from preceding and following signal portions. The weaker version was considered more realistic because human listeners clearly have the ability to combine multiple sources of information and will make use of that ability whenever multiple sources are available. Peripheral auditory processes do not seem to have the integrative power to combine temporally distributed phonetic information. On the contrary, it was argued that adaptation helps differentiate the signal into contrasting auditory components.

From a review of the physiological and psychoacoustic literature it was concluded that short-term adaptation almost certainly does take place in the human auditory system during speech perception. The internal representation of the auditory signal from which phonetic information is derived, particularly at points following rapid spectral change, is therefore different from the one visible in a spectrogram or oscillogram. However, does adaptation have any consequences for the intelligibility of speech? Summerfield et al. (1984) have pointed out some putative general advantages, such as improvement of the signal-noise ratio, but such advantages exist only relative to a hypothetical auditory system or speech recognition device in which no adaptation occurs.

The former may not exist, since adaptation may well be a general design feature of neural systems. As to the latter, it should be noted that adaptation can only enhance existing spectral change, not create it. Its perceptual effect is thus comparable to a lowering of the threshold for spectral change detection on an arbitrary scale, which a machine can easily emulate, and whose net effect is zero. Thus, there is perhaps no real "advantage" to be had from adaptation and spectral enhancement, except perhaps when the spectral change is right at the detection threshold. Similar conclusions have been drawn from studies of the effects of bandwidth narrowing and spectral enhancement on speech intelligibility in the hearing-impaired (Leek, Dorman, & Summerfield, 1987; Summerfield, Foster, & Tyler, 1985).

It is still meaningful, however, to ask whether any perceptual *disadvantage* results from a reduction of adaptation, achieved by stimulus manipulations in the laboratory. The problem here is that such manipulations may have repercussions at all levels of the system, so it is not clear whether a performance decrement results specifically from the absence of peripheral spectral enhancement or from interference with a more central process of spectral comparison or integration. This problem beset Experiments 1-3, in which auditory short-term adaptation was interfered with and identification performance decreased accordingly. Had it not decreased at all, this would have been evidence that adaptation plays no role in the perception of prevocalic nasal consonants. As it was, the only indication that adaptation is perhaps unimportant was the rather small decrease in intelligibility consequent upon temporal separation of murmur and vowel portions (Experiment 3).

Experiment 4 added two other relevant findings. Reduction of murmur duration, which presumably diminished the degree of adaptation, caused a performance decrement, but only at the very shortest duration. Although a ceiling effect may have imposed some limits, this finding is somewhat unfavorable to the adaptation hypothesis. The other finding was that mismatched murmurs did not lead to a performance decrement in [-*a*] and [-u] syllables, which confirmed a prediction of the adaptation hypothesis. A very different result was obtained with [-i] syllables, however, which was more difficult (but not impossible) to reconcile with the adaptation hypothesis. All in all, the hypothesis emerged relatively unscathed from Experiments 1-4.

Experiment 5 considered two alternative hypotheses, neither of which received much support. First, place of articulation perception was no more accurate for stimuli whose nasal manner was correctly perceived. Second, smoothing the abrupt stimulus onset caused by removal of the murmur engendered only a small improvement in identification performance—not enough to account for the high intelligibility of combined murmur and vowel onset cues.

The adaptation hypothesis was still viable at this point. Experiments 6 and 7, however, yielded results that were clearly contrary to its predictions: A simulation of spectral enhancement at the onset of isolated vowel portions generally harmed, rather than improved, place of articulation identification. It may be argued that the situation was too artificial, and that spectral change information can be utilized only when the signal portion preceding the change (the murmur) is physically present. This objection, however, would be tantamount to saying that spectral change information is obtained by a more central computational process, rather than by peripheral adaptation. Or, in other words, it is the spectral change itself that is perceptually important, and not its auditory transformation through adaptation.

To compute the relationship between two stimulus components, it seems necessary that relatively analog representations of these components be available to the central nervous system. Once the murmur has been processed separately and encoded as a vector of categorical possibilities (Chistovich, 1985; Massaro & Oden, 1980), there is no way of recovering spectral relationships during processing of the vowel. This consideration points to *auditory memory* as a mediator in the central perceptual integration of stimulus components. That is, listeners may be able to hold on to a relatively faithful auditory representation of the nasal murmur even across a stretch of intervening noise or silence, and to compare that memory trace to the vowel onset spectrum. Moreover, even though the temporal separations employed in the present experiments are within the range of short-term auditory storage (Cowan, 1984), it seems likely that listeners rely on long-term auditory storage in making spectral comparisons, one reason being that the vowel would tend to "overwrite" the murmur in a sensory buffer (Cowan, 1984). Long-term auditory storage may last for a number of seconds, depending on the amount of detail to be retained. Even a life span of one second would be more than sufficient to account for the findings of the present study. This explanation is consistent with the very gradual decline in performance as a function of temporal separation.

Why are the murmur and vowel components integrated at all? The auditory adaptation hypothesis advanced by Kurowski and Blumstein (1984) was an attempt to provide a low-level explanation: Integration is assumed to occur because of general principles of auditory processing, and the speech perceiver merely needs to "pick up" the neatly parceled, unitary auditory properties to arrive at phonetic judgments. It seems, however, that auditory operations alone are insufficient to account for the perceptual integration of speech components. Indeed, it is not the signal portions themselves that are integrated (i.e., they remain audible as separate auditory events; this is even more obvious in the case of fricative-vowel syllables, for example) but the *information* they convey. The information, to deserve that name, must inform the listener about some event he or she has learned (or was born) to recognize. The rationale for information integration thus must be sought in the listener's mental representations of common speech patterns, which in turn reflect the regular occurrences of acoustic (and articulatory) events in speech production (see also Repp, 1987a, 1987b). That is, the cues provided by the nasal murmur and by the following vowel are "integrated" because they, and their relationship (i.e., the pattern of spectral change reflecting articulatory movement), all contribute information about place of articulation of prevocalic consonants, and because listeners *know* this from long experience with speech as individuals and as members of the human species. In other words, the perceptual integration of the articulatory information conveyed by auditorily distinct speech components is a centrally guided, not a peripheral phenomenon. It reflects the listener's knowledge of the way speech is patterned, not principles governing the operation of the auditory system.

## References

Abbas, P. J., & Gorga, M. P. (1981). AP responses in forward-masking paradigms and their relationship to responses of auditory-nerve fibers. *Journal of the Acoustical Society of America, 69*, 492-499.

Abramson, A. S., Nye, P. W., Henderson, J. B., & Marshall, C. W. (1981). Vowel height and the perception of consonantal nasality. *Journal of the Acoustical Society of America, 70*, 329-339.

Bell-Berti, F., Baer, T., Harris, K. S., & Niimi, S. (1979). Coarticulatory effects of vowel quality on velar function. *Phonetica, 36*, 187-193.

Carden, G., Levitt, A., Jusczyk, P. W., & Walley, A. (1981). Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception & Psychophysics, 29,* 26-36.

Chistovich, L. A. (1985). Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America, 77,* 789-805.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96,* 341-370.

Danaher, E. M., & Pickett, J. M. (1975). Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss. *Journal of Speech and Hearing Research, 18,* 261-271.

Delgutte, B. (1980). Representation of speech-like sounds in the discharge patterns of auditory nerve-fibers. *Journal of the Acoustical Society of America, 68,* 843-857.

Delgutte, B., & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America, 75,* 897-907.

Eggermont, J. J. (1985). Peripheral auditory adaptation and fatigue: A model oriented review. *Hearing Research, 18,* 57-71.

Furui, S. (1986). On the role of spectral transition for speech perception. *Journal of the Acoustical Society of America, 80,* 1016-1025.

Green, D. M. (1983). Profile analysis: A different view of auditory intensity discrimination. *American Psychologist, 38,* 133-142.

Hannley, M., & Dorman, M. F. (1983). Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss. *Journal of the Acoustical Society of America, 74,* 40-51.

Harris, D. M., & Dallos, P. (1979). Forward masking of auditory nerve fiber responses. *Journal of Neurophysiology, 42,* 1083-1107.

Kurowski, K., & Blumstein, S. E. (1984). Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants. *Journal of the Acoustical Society of America, 76,* 383-390.

Lahiri, A., Gewirth, L., & Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants. *Journal of the Acoustical Society of America, 76,* 391-404.

Larkey, L. S., Wald, J., & Strange, W. (1978). Perception of synthetic nasal consonants in initial and final syllable position. *Perception & Psychophysics, 23,* 299-312.

Leek, M. R., Dorman, M. F., & Summerfield, Q. (1987). Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 81,* 148-154.

Malécot, A. (1956). Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique. *Language, 32,* 274-284.

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America, 67,* 996-1013.

Miller, J. L. (1977). Nonindependence of feature processing in initial consonants. *Journal of Speech and Hearing Research, 20,* 519-528.

Ohde, R. N., & Sharf, D. J. (1981). Stop identification from vocalic transition plus vowel segments of CV and VC syllables: A follow-up study. *Journal of the Acoustical Society of America, 69,* 297-300.

Parker, E. M., & Diehl, R. L. (1984). Identifying vowels in CVC syllables: Effects of inserting silence and noise. *Perception & Psychophysics, 36,* 369-380.

Plomp, R. (1964). Rate of decay of auditory sensation. *Journal of the Acoustical Society of America, 36,* 277-282.

Pols, L. C. W., & Schouten, M. E. H. (1978). Identification of deleted consonants. *Journal of the Acoustical Society of America, 64,* 1333-1337.

Pols, L. C. W., & Schouten, M. E. H. (1981). Identification of deleted plosives: The effect of adding noise or applying a time window (A reply to Ohde and Sharf). *Journal of the Acoustical Society of America, 69,* 301-303.

Recasens, D. (1983). Place cues for nasal consonants with special reference to Catalan. *Journal of the Acoustical Society of America, 73,* 1346-1353.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92,* 81-110.

Repp, B. H. (1986). Perception of the [m]-[n] distinction in CV syllables. *Journal of the Acoustical Society of America, 79,* 1987-1999.

Repp, B. H. (1987a). The role of psychophysics in understanding speech perception. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception.* The Hague: Martinus Nijhoff.

Repp, B. H. (1987b). Integration and segregation in speech perception. In M. Remmel (Ed.), *Proceedings of the Twelfth International Congress of Phonetic Sciences.* Tallinn, Estonia, USSR.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 44,* 1735-1736.

Smith, R. L. (1979). Adaptation, saturation, and physiological masking in single auditory nerve fibers. *Journal of the Acoustical Society of America, 65,* 166-178.

Summerfield, Q., & Assmann, P. (1987). Auditory enhancement and speech perception. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception.* The Hague: Martinus Nijhoff.

Summerfield, Q., Foster, J., & Tyler, R. (1985). Influences of formant bandwidth and auditory frequency selectivity on identification of place of articulation in stop consonants. *Speech Communication, 4,* 213-229.

Summerfield, Q., Haggard, M., Foster, J., & Gray, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. *Perception & Psychophysics, 35,* 203-213.

Viemeister, N. F., & Bacon, S. P. (1982). Forward masking by enhanced components in harmonic complexes. *Journal of the Acoustical Society of America, 71,* 1502-1507.

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin, 96,* 371-383.

Westerman, L. A., & Smith, R. L. (1984). Rapid and short-term adaptation in auditory nerve responses. *Hearing Research, 15,* 249-260.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics, 35,* 49-64.

Whalen, D. H., & Samuel, A. G. (1985). Phonetic information is integrated across intervening nonlinguistic sounds. *Perception & Psychophysics, 37,* 579-587.

Widin, G. P., & Viemeister, N. F. (1979). Intensive and temporal effects in pure-tone forward masking. *Journal of the Acoustical Society of America, 66,* 388-395.

Wilson, J. P. (1970). An auditory after image. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing.* Leiden: A. W. Sijthoff.

Zwislocki, J., Pirodda, E., & Rubin, H. (1959). On some poststimulatory effects at the threshold of audibility. *Journal of the Acoustical Society of America, 31,* 9-14.

# DIFFERENCE IN SECOND-FORMANT TRANSITIONS BETWEEN ASPIRATED AND UNASPIRATED STOP CONSONANTS PRECEDING [*a*]*

**Bruno H. Repp and Hwei-Bing Lin** †

**Abstract**. *Perceptual experiments with synthetic speech have shown that the category boundary on an acoustic [pa]-[ta] (/ba/-/da/) continuum (obtained by varying the onset frequencies of the second and third formants) is closer to the labial endpoint than the boundary on a [pʰa]-[tʰa] (/pa/-/ta/) continuum. Of several possible explanations, the most plausible seems to be that natural unaspirated and aspirated stops have different formant transitions. To supplement limited data on this point in the literature, we conducted an acoustic analysis of CV syllables produced by 10 male speakers of American English. The results show very clearly that the second formants of [pʰa] and [tʰa] start 100-200 Hz higher than those of [pa] and [ta] and reach comparable frequency values only at voicing onset. This difference, which is probably an acoustic consequence of subglottal coupling during aspiration, seems to be part of a listener's tacit knowledge of phonetic regularities and thus explains the perceptual boundary shift. It also needs to be taken into account in realistic speech synthesis.*

## Introduction

A highly reliable finding of perceptual studies using synthetic CV syllables forming place of articulation continua is that the category boundary on an unaspirated [p*a*]-[t*a*] (i.e., /b*a*/-/d*a*/) continuum is closer to the labial endpoint than the corresponding boundary on an aspirated [pʰ*a*]-[tʰ*a*] (i.e., /p*a*/-/t*a*/) continuum (Alfonso & Daniloff, 1980; Massaro & Oden, 1980; Miller, 1977; Oden & Massaro, 1978; Ohde & Stevens, 1983; Repp, 1978). In each of these studies, the stimuli in the two continua differed in the onset frequencies and transitions of the second and third formants ($F_2$ and $F_3$), whereas the difference between the two continua rested on voice

---

* *Language and Speech*, in press.

† Also Department of Linguistics, University of Connecticut, Storrs, CT 06268

*Acknowledgment*. This research was supported by NICHD Grant HD-01994 to Haskins Laboratories. We are grateful to Ignatius Mattingly, Richard McGowan, Susan Nittrouer, and especially Kevin Munhall for their excellent comments which helped improve this paper. A short version of the paper was presented by the second author at the 113th Meeting of the Acoustical Society of America in Indianapolis, May 1987.

onset time (VOT). In the case of aspirated stops, this meant a delay in voicing onset, presence of aspiration noise, and attenuation or complete suppression of the first formant ($F_1$) during the aspirated interval. Formant transitions and VOT thus were varied in a strictly orthogonal fashion.

No satisfactory explanation has been provided for the perceptual boundary shift, although several authors have speculated about its causes. If we include several additional possibilities that have occurred to us, no less than six different hypotheses result, which we shall discuss briefly to show that all but the one addressed by our study (No. 6) are unlikely candidates.

(1) *Feature processing interaction.* Miller (1977) attributed the boundary shift to nonindependence in phonetic feature processing. (See also Haggard, 1970; Oden & Massaro, 1978; Sawusch & Pisoni, 1974; Smith, 1973.) At the time, when feature detector theory was at the height of its popularity (see Remez, 1987), this hypothesis may have seemed to have some explanatory value. Basically, however, it is just a restatement of the finding, since it would be just as valid if the boundary shift went in the opposite direction. One testable prediction may be derived from this hypothesis, however: The shift in the place of articulation boundary should be a step function of VOT; that is, for a series of place of articulation continua differing by small increments in VOT, the perceptual boundary between labial and alveolar categories should change abruptly as VOT crosses the phonological voicing boundary but should remain relatively constant within voicing categories. In other words, the location of the place boundary should be a function of the perceived voicing category (the discrete response of a hypothetical "voicing detector"), not of VOT. In several experiments using appropriate stimulus arrays, Oden and Massaro (1978) and Massaro and Oden (1980) actually obtained results consistent with this prediction, although they nevertheless chose to emphasize the "relatively continuous" nature of the boundary change (Massaro & Oden, 1980, p. 1003). Repp (1978), on the other hand, obtained fairly continuous place boundary changes as a function of VOT; however, VOT varied over a smaller range in his stimuli. In view of these inconclusive data, the feature processing interaction hypothesis cannot be dismissed, but it has little explanatory power in the context of contemporary theorizing, especially since it is indifferent to the direction of the boundary shift. The same can be said about Oden and Massaro's (1978) feature integration model, which, even though it assumes independent processing of acoustic features, represents the phonetic feature interaction at the level of mental category prototypes. The model fits the data well, but it does not explain the direction of the effect.

(2) *Presence versus absence of* $F_1$. A second hypothesis is that the boundary shift originates in the auditory system: Some auditory interaction may make the $F_2$ and $F_3$ transitions of aspirated stops appear to be lower in frequency than those of unaspirated stops, or may increase the relative perceptual salience of rising (labial) versus falling (alveolar) $F_2$ and $F_3$ transitions in aspirated as compared to unaspirated stops. The first formant could be involved in such an interaction. Because $F_1$ tends to be weak during natural aspiration, and because "$F_1$ cutback" is in fact an important cue for phonological voicelessness in initial English stop consonants (Liberman, Delattre, & Cooper, 1958), $F_1$ has been attenuated as a matter of routine in the synthesis of aspirated stop consonants. There is also evidence in the literature that, in certain situations, the $F_1$ transition, when it is present, may influence the perception of transitions in the higher formants: When a syllable is split between the ears, so that $F_1$ goes to one ear and $F_2$ to the other ear, the discriminability of $F_2$ transitions is improved relative to a monaural or binaural condition (Danaher & Pickett, 1975; Rand, 1974). This improvement has been attributed to a release from peripheral "upward spread of masking" by $F_1$. It seems reasonable that such

masking would have a greater effect on $F_2$ transitions that are close in frequency to $F_1$ and/or have a similar (rising) trajectory; thus it might decrease the relative salience of labial transitions in unaspirated stops, so attenuation of $F_1$ in aspirated stops would then result in a relative enhancement of these transitions, in accord with the observed perceptual boundary shift. In dichotic split-formant studies, Perl and Haggard (1974) and especially Perl (1975) did observe "a tendency for increased dichotic release from masking where initial $F_2$ transitions tend towards the same slope as accompanying $F_1$ transitions" (Perl, 1975, p. 36). Unfortunately, most other relevant studies failed to show such trends (Grunke & Pisoni, 1982; Hannley & Dorman, 1983; Nusbaum, Schwab, & Sawusch, 1983; Schwab, 1981; Turek, Dorman, Franks, & Summerfield, 1980). In addition, informal observations by the first author suggest that synthetic syllables in which phonological voicelessness is cued solely by $F_1$ cutback without accompanying aspiration noise (cf. Liberman et al., 1958) do not exhibit any place boundary shift. The upward spread of masking hypothesis thus seems untenable.

(3) *Absence of release burst.* A third possible explanation takes note of the fact that most studies have employed synthetic syllables without release bursts. Alveolar release bursts, because of their different spectral energy distribution, are more intense than labial release bursts, and aspirated stops tend to have stronger bursts than unaspirated stops (Zue, 1976). Burst amplitude (with spectral properties held constant) has been shown to be a secondary place of articulation cue: Listeners report more labial stop percepts when the amplitude is low than when it is high (Ohde & Stevens, 1983; Repp, 1984). Thus, if listeners *expect* a burst, its absence may lead to a general bias toward labial stop percepts, and this bias may be larger for stimuli that normally have stronger release bursts, viz., aspirated stops. In other words, the absence of a strong burst may make a stimulus sound even more labial than does the absence of a weak burst. However, Ohde and Stevens (1983) employed aspirated and unaspirated stimuli that included synthetic bursts and still found a large place boundary shift as a function of aspiration. Therefore, the "missing burst" hypothesis seems less promising now than it did a few years ago. Besides, it is almost impossible to test rigorously because of the difficulty of synthesizing release bursts that are both realistic and matched to the formant transitions on a place of articulation continuum.

(4) *VOT as a place cue.* It is well known that alveolar stops have longer VOTs than labial stops, especially in their aspirated forms, although the difference is not very large and there is substantial overlap of the VOT distributions (see, e.g., Lisker & Abramson, 1967; Ohde, 1984). Even so, it is conceivable that the temporal aspect of VOT serves as a weak place cue in aspirated stops, such that listeners are somewhat more likely to perceive labials when VOT is relatively short, and alveolars when VOT is relatively long. If the VOTs of the synthetic [p$^h$*a*]-[t$^h$*a*] stimuli used in earlier studies were on the short side, the place boundary shift in favor of labial responses could be accounted for. The longest VOT used by Oden and Massaro (1978) and Massaro and Oden (1980) was 40 ms; that employed by Repp (1978) was 42 ms; Miller (1977) and Ohde and Stevens (1983) used a VOT of 50 ms for their aspirated stops; and Alfonso and Daniloff (1980) used a VOT of 60 ms. The average VOT of [p$^h$*a*] and [t$^h$*a*] produced in isolation is about 70 ms, with the VOT of [t$^h$*a*] being some 10 ms longer than that of [p$^h$*a*] (Lisker & Abramson, 1967; present study). Thus all VOTs used in previous synthesis were indeed on the short (labial) side. It is noteworthy, however, that the largest place boundary shifts (about 145 Hz in terms of $F_2$ onset frequency) were obtained by Alfonso and Daniloff (1980), who used the longest VOT for their aspirated continuum. This observation, together with the great variability of VOTs in natural speech, makes it unlikely that VOT could be responsible for the boundary shift.

(5) *Aspiration noise spectrum and/or intensity as a place cue.* Massaro and Oden (1980) proposed that the aspiration noise itself may provide a cue for labial place of articulation (see also Ohde & Stevens, 1983). At first glance, this hypothesis seems to ignore the fact that in synthetic stimuli (as in natural speech) the aperiodic source passes through the same $F_2$ and $F_3$ filters as the periodic source, leading to similar spectral shapes above $F_1$. It is possible, however, that differences in the spectral slope and/or amplitude of periodic and aperiodic source spectra somehow contribute to the perceptual boundary shift, especially if they deviate from what is observed in natural speech. Unfortunately, these parameters are commonly omitted from descriptions of synthetic stimuli, and information about their magnitudes in natural speech is also hard to come by. Massaro and Oden did find that labial responses increased further when aspiration noise intensity was increased; however, since labial responses increased with VOT (up to 40 ms, the longest value used) in their study, the result may reflect the fact that stimuli with higher aspiration levels are phonetically equivalent to stimuli with longer VOTs, perhaps due to a time-intensity reciprocity in auditory perception (Darwin & Seton, 1983; Repp, 1979). Certainly there is no reason to believe that natural labial stops are characterized by more intense aspiration than alveolar stops. In summary, while the global acoustic characteristics of natural aspiration bear closer examination, it seems unlikely that they vary with place of articulation and, hence, that they could function as secondary place of articulation cues.

(6) *Different formant transitions in unaspirated and aspirated stops.* The sixth and final hypothesis is that the formant transitions are different in aspirated and unaspirated stops, so that listeners apply different criteria for place decisions along a formant transition continuum depending on whether aspiration is present or absent. Despite a long tradition of synthesizing unaspirated and aspirated stops with identical formant transitions for use in perceptual experiments (which may derive, in part, from the "locus" theory of Delattre, Liberman, & Cooper, 1955), there is in fact some limited support for this hypothesis in the acoustic phonetics literature. Fant (1973) reports that /p/ (i.e., [$\mathrm{p}^h$]) tends to have higher $F_2$ onsets than /b/ (i.e., [p]) before back vowels such as /a/. However, his very limited data derive from a single speaker of Swedish, and some of the formant frequencies reported seem unusually low. Similar data for English collected by Lehiste and Peterson (1961) and replotted by Fant (1973) are suggestive at best. More convincing are Gay's (1978) spectrographic measurements of $F_2$ onset frequencies in syllables produced by three male American speakers: $F_2$ onset in /pap/ and /pup/ was about 180 Hz higher than in /bap/ and /bup/; however, it was about 125 Hz lower in /pip/ than in /bip/.

Gay mentions three possible causes of the difference in formant transitions preceding back vowels: (a) The *coarticulatory hypothesis*: Fant (1973) speculated that /b/ is coarticulated more strongly with a following back vowel (i.e., the tongue is more nearly in position for the vowel before the release of the stop closure) than is /p/, while no such difference exists between /d/ and /t/. (b) The *release timing hypothesis*: As the articulators begin to move towards the vowel, the release of aspirated stops may occur earlier in time than that of unaspirated stops, so that energy begins while the articulators are still farther away from the vowel target (Öhman, 1965; see Fant, 1973, p. 118). The acoustic consequences are similar to those predicted by the coarticulatory hypothesis, but it should be possible to overlay the formant trajectories of aspirated and unaspirated stops after correcting for the time shift (cf. Fant, 1973). (c) The *subglottal coupling hypothesis*: The higher $F_2$ onsets for aspirated stops may arise from the open glottis during aspiration. This acoustic explanation appears very plausible in view of research by Lehiste (1964, cited in Lehiste, 1970) and Kallail and Emanuel (1984a, 1984b) on whispered vowels, in which especially $F_1$ but also $F_2$ and $F_3$ tend to be higher than in phonated vowels, with the possible exception of high front vowels. Indeed, glottal opening is likely to be wider at the beginning of aspiration than during whisper (Catford, 1977). Fant, Ishizaka, Lindqvist, and Sundberg (1972) have modeled these effects of subglottal coupling, which may include additional subglottal formants in the aspiration spectrum, especially right after the release.

A clear demonstration of higher formant frequencies (especially of $F_2$) in aspirated than in unaspirated stop consonants preceding [*a*] would be of value for three reasons: First, the relevant data in the literature are incomplete and not easy to find; in particular, there have been no comparisons of the complete formant transitions in unaspirated and aspirated stops for both labial and alveolar places of articulation. Second, such data would provide an important guideline for realistic speech synthesis. Third, they would provide a sufficient explanation of the perceptual boundary shift and provide yet another illustration that listeners engaging in linguistic classification rely on tacit knowledge of a wealth of phonetic detail (see Repp, 1987).

Only the syllables [p*a*], [t*a*], [p$^h$*a*], [t$^h$*a*], were considered in this study, because they were the endpoints of the continua used in previous perceptual studies. Nevertheless, it was possible even in this limited context to address the three hypotheses about the origin of differences in formant frequencies between unaspirated and aspirated stops, if any were found: (a) If Fant's coarticulatory hypothesis is correct, the difference should be more pronounced for labial than for alveolar stops, since the tongue body is less free to anticipate the shape of the following vowel during alveolar closure. Also, the time course of the labial $F_2$ transition should be independent of VOT in aspirated tokens; that is, it should be a function of the movements of the upper articulators only. (b) If Öhman's release timing hypothesis is correct, the results should be similar, but in addition it should be possible to superimpose the average formant tracks of unaspirated and aspirated tokens by shifting them in time relative to each other. Thus, a finding of rising $F_2$ transitions for [p*a*] but falling $F_2$ transitions for [p$^h$*a*] would be incompatible with the release timing hypothesis, but not necessarily with Fant's coarticulatory hypothesis. (c) If the subglottal coupling hypothesis is correct, the $F_2$ difference between aspirated and unaspirated stops should be present for both labial and alveolar stops and should disappear with voicing onset in aspirated tokens. Of course, these hypotheses are not mutually exclusive, and more than one explanation may be supported by the data.

In addition to providing measurements of $F_2$ trajectories to address these principal hypotheses, the present study also yielded data on $F_1$ and $F_3$ frequencies, and on the spectral tilt and relative amplitude of aspiration—information that is difficult to locate in the literature but is useful for speech synthesis.

## Methods

Ten male speakers of American English produced the syllables [p*a*], [t*a*], [p$^h$*a*], [t$^h$*a*], five times in random order, reading from a list of randomized syllables spelled BA, DA, PA, TA. They were recorded in a sound-insulated booth using a Sennheiser microphone and an Otari MX5050 tape recorder located in an adjacent booth. The mouth-to-microphone distance was about 20 inches. All 200 utterances were low-pass filtered at 4.9 kHz and digitized at a sampling rate of 10 kHz with high-frequency pre-emphasis. Each file was edited to eliminate silence or (rare) voicing preceding the release. A 14-coefficient LPC analysis was then conducted using a 20-ms Hamming window advancing in 10-ms steps, and formant frequencies were estimated using the root solving method (ILS package, Version 4.0, distributed by Signal Technology, Inc.).

The resulting arrays of formant frequencies as a function of time were cleaned up by hand to eliminate occasional spurious peaks, to make sure that all frequencies were aligned with the appropriate formants, and to deal with the problem of missing values. One speaker was excluded from further analysis because of insufficient $F_2$ data for labial stops. For the other speakers, missing formant frequencies were filled in by interpolating between preceding and following values or, if they occurred at the onset, by extending the first existing value backward in time. Missing frequencies were especially common in the initial time frames; this was not surprising, since release bursts often do not have a clear formant structure. Thirty-eight percent of the $F_2$ data were missing in frame 1, 19% in frame 2, and from 12% to 3% in frames 3-10. Eighty-six percent of all missing values were in aspirated tokens; of these, 62 percent were in [p$^h$a] tokens and 38 percent in [t$^h$a] tokens. For $F_3$, the percentages of filled-in values were 28% in frame 1 and between 7 and 15% in frames 2-10. While interpolation of missing $F_2$ and $F_3$ values in later frames should not have distorted the analysis results in any way, the filling in of missing initial values by level extension of later values (a conservative procedure) may have resulted in an underestimation of existing differences in formant frequencies between unaspirated and aspirated tokens at onset. $F_1$, of course, was generally absent during aspiration and was also spurious in unaspirated tokens for two speakers. To compare $F_1$ in unaspirated labials and alveolars, the $F_1$ data of the eight speakers with fairly complete values were analyzed after filling in missing values (36% in frame 1, 2-6% in frames 2-10).

Voice onset times of aspirated tokens were measured in a waveform display by locating the onset of the first glottal pulse. In addition, to corroborate the LPC analysis results and to examine the spectral and amplitude characteristics of aspiration, FFT spectra of all utterances were obtained from 20-ms Hamming windows centered 10, 30, and 50 ms after the release. To reduce random level fluctuations, the spectra were averaged over the five repetitions of each syllable by each speaker. From these average spectra we picked $F_2$ peaks by eye wherever possible, interpolating if there were two closely adjacent peaks in the relevant region. This yielded complete estimates of $F_2$ frequencies for all 10 speakers at the three time points for [pa] and [ta]; for [t$^h$a], only 2 data points (7% of the data) were missing; for [p$^h$a], however, peaks could not be located in 10 instances (33% of the data). As with the LPC data, the missing values were interpolated or extrapolated from the existing ones, so as to have a complete matrix for calculation of means and for statistical analysis.

## Results and Discussion

### *$F_2$ Transitions*

Because of considerable differences in utterance durations for different speakers, only the first 110 ms of each token (i.e., 10 overlapping 20-ms analysis time frames) were considered. The cleaned-up arrays of formant values were averaged across all tokens of all speakers to obtain an overall picture of the differences in formant transitions. These average $F_2$ transitions are plotted as the connected points in Figure 1. It is evident that both aspirated syllable types had higher $F_2$ onsets than their unaspirated counterparts, and that this difference gradually decreased over the first 70 ms or so. Right after the release the difference was larger for labials than for alveolars, but after 30 ms it seemed independent of place of articulation. In addition, it may be noted that

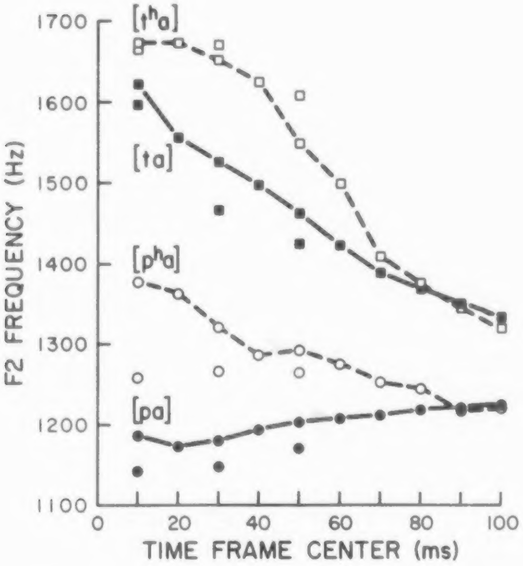

**Figure 1.** The connected points show the average second formant ($F_2$) transitions over the first 100 ms of [*pa*], [p$^h$*a*], [*ta*], and [t$^h$*a*], as determined by LPC analysis. Each transition represents the average of 45 utterances (5 tokens from each of 9 speakers). The unconnected points represent average $F_2$ frequency estimates from FFT analysis of all 10 speakers' productions. Formant frequencies are plotted at the centers of the 20 ms time windows.

$F_2$ was higher for alveolar than labial tokens well beyond the first 100 ms. Formant transitions thus may be a good deal longer than the (approximately) 50 ms often cited in the literature and employed in speech synthesis.

A repeated-measures analysis of variance was conducted on the token averages with place of articulation, aspiration, and time as factors. All main effects and interactions were significant at $p = .0005$ or less, except for the place by aspiration interaction, which was nonsignificant. The overall magnitude of the aspiration effect was thus similar for labial and alveolar stops. The triple interaction, $F(9,72) = 5.71, p < .0001$, however, confirms that the aspiration effect was smaller for alveolar than for labial stops immediately after the release. Separate analyses of labial and alveolar tokens showed that the unaspirated/aspirated difference was significant for both places of articulation—labial: $F(1,8) = 32.67, p = .0004$; alveolar: $F(1,8) = 15.03, p = .0047$. In addition, their decrease as a function of time was reflected in highly significant interactions between aspiration and time—labial: $F(9,72) = 29.69, p < .0001$; alveolar: $F(9,72) = 11.51, p < .0001$. This pattern was shown by all individual speakers.

Similar analyses were conducted on the $F_2$ frequency estimates derived from FFT spectra; the averages are plotted as the unconnected points in Figure 1. As pointed out in the Methods section, the data for [p$^h$*a*] were somewhat unreliable, which explains the major discrepancy between the LPC and FFT frequency estimates for that syllable. For the other syllables, there was reasonable agreement between the two sets of data, although FFT estimates seemed to be systematically lower than LPC estimates for unaspirated stops. Absolute differences aside, the FFT data clearly corroborate the finding of higher $F_2$ frequencies during aspiration. In the overall analysis of variance, all effects except the place by aspiration interaction were significant at $p = .01$ or less. Tested separately, the main effect of aspiration was significant for both labial, $F(1,9) = 7.72, p =$

.0214, and alveolar stops, $F(1,9) = 34.09, p = .0002$; for the latter there was also a significant change of the effect over time, $F(2,18) = 8.27, p = .0028$.

The magnitude of the difference for labials at release is in good agreement with Gay's (1978) data, as are the absolute LPC-derived formant frequencies. The magnitude of $F_2$ difference between phonated and whispered [*a*] reported by Kallail and Emanuel (1984b) is also similar. This last observation, together with the finding of similar differences for labials and alveolars, except right after the release, suggests that the explanation is to be found in the open glottis during aspiration.

Of the two alternative explanations, Öhman's release timing hypothesis seems to be inconsistent with the present data. Even granting possible distortions due to averaging over tokens representing different vocal tract sizes and speaking rates, there is no way the transitions for unaspirated and aspirated tokens could be time-shifted to coincide in Figure 1. This is especially true in the case of [p*a*], which has a barely rising $F_2$ transition, and [p$^h$*a*], which has a clearly falling one. Thus, this hypothesis can be dismissed. Fant's coarticulation hypothesis predicted a smaller difference for alveolars than for labials, which was found immediately after the release but not some tens of milliseconds later. It is possible that, as the tongue is freed from the constraint of the alveolar closure, it rapidly adjusts to the following vowel shape, and more so in [t*a*] than in [t$^h$*a*]. (Alternatively, the presence of a frication source at the alveolar constriction may obscure any existing $F_2$ differences during alveolar release bursts.) The coarticulatory hypothesis thus is not incompatible with the data in Figure 1, even though Fant himself commented only on labial stops.

Another prediction of Fant's hypothesis, however, is that the time course of the $F_2$ difference should be independent of when voicing starts in aspirated tokens. The subglottal coupling hypothesis, on the other hand, predicts that the difference should end at voicing onset. The $F_2$ trajectories for [p$^h$*a*] and [t$^h$*a*] shown in Figure 1 were obtained by averaging over aspirated tokens with VOTs ranging from 40 to 126 ms, with an average of 70 ms (66 ms for labials, 73 ms for alveolars), which resulted in considerable smearing in the time domain. An alternative way to analyze the data is to line up all aspirated tokens at voice onset rather than at the release. Figure 2 shows the average $F_2$ frequencies in the vicinity of voice onset after lining up aspirated tokens in this way, with unaspirated tokens lined up correspondingly by yoking them to aspirated tokens of the same speaker and shifting them by the same amount along the time axis. It can be seen that the $F_2$ difference indeed disappears at voice onset for alveolar stops, and nearly so for labial stops. In analyses of variance on the five time frames following voice onset, no significant $F_2$ differences were obtained for either labials or alveolars. An additional analysis including rank-ordered VOT as a factor was conducted to determine whether $F_2$ onset frequency in aspirated stops increased with VOT. The result was negative.

Had the differences in $F_2$ trajectories extended beyond voicing onset or had they ended much sooner, the coarticulatory hypothesis might have been favored over the subglottal coupling hypothesis. On the other hand, a positive correlation between $F_2$ onset frequency and VOT in aspirated stops would have supported the latter hypothesis. As it is, the data are still consistent with both hypotheses, though the subglottal coupling hypothesis would seem to provide a more parsimonious account: The acoustic consequences of subglottal coupling are necessary effects, while differences in the position of the upper articulators are not (as long as no direct observations of articulation show they do exist). The gradual decline in the $F_2$ difference prior to voice onset

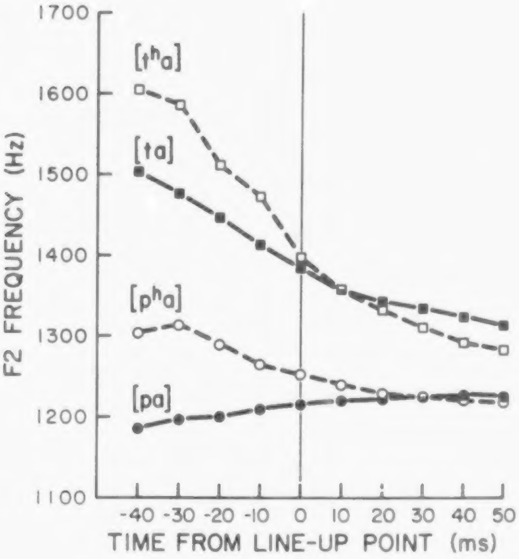

**Figure 2.** Average second formant ($F_2$) frequencies in the vicinity of voicing onset for [p$^h a$] and [t$^h a$] tokens lined up at voicing onset, and for yoked [p$a$] and [t$a$] tokens lined up at corresponding time points.

probably reflects the gradual narrowing of the glottal opening before voicing starts (see, e.g., Hirose, 1977; Kagaya, 1974). The smaller difference between $F_2$ of [t$a$] and [t$^h a$] right after release may be due to broadband frication noise generated while the constriction is narrow. Subglottal coupling thus provides a sufficient explanation of the observed differences in $F_2$ trajectories.

### *$F_1$ and $F_3$ Transitions*

We also examined differences in $F_3$ transitions in the same manner. However, there were no significant $F_3$ differences as a function of aspiration in either labials or alveolars, whether aligned at release or at voice onset. Kallail and Emanuel (1984b), too, found only a very small (presumably nonsignificant) $F_3$ difference between voiced and whispered [$a$].

$F_1$, on the other hand, is strongly affected by a change in source, being about 250 Hz higher in whispered than in phonated male [$a$] (Kallail & Emanuel, 1984b), but its increased bandwidth makes frequency measurements difficult, and we did not attempt to determine $F_1$ frequencies during aspiration. We did compare $F_1$ transitions in unaspirated [p$a$] and [t$a$] for eight subjects (for two subjects the LPC analysis did not yield reliable $F_1$ estimates, but the subject excluded from the $F_2$ analysis was included here) and found a significant difference, $F(1,7) = 21.87, p = .0023$, which decreased over time, $F(9,63) = 17.33, p < .0001$. All subjects showed higher $F_1$ onsets in [p$a$] than in [t$a$]; the averages were 669 and 589 Hz, respectively. After 100 ms, this 80 Hz difference had dwindled to 28 Hz.

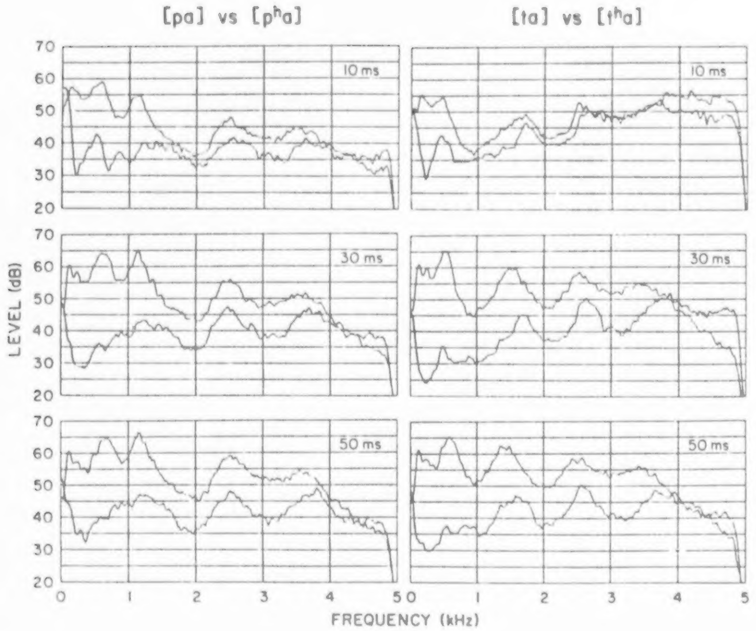

**Figure 3.** Average Fourier (FFT) spectra of unaspirated and aspirated stops at three points in time, calculated using Hamming windows centered 10, 30, and 50 ms after the release. Each spectrum represents the average of 50 utterances (5 tokens from each of 10 speakers). The upper function in each panel represents the unaspirated stops, and the lower function the aspirated ones. All spectra include high-frequency pre-emphasis of approximately 6 dB/octave above 1 kHz, and less below.

### *Aspiration Noise: Spectral Tilt and Relative Amplitude*

Finally, we compared spectral cross-sections of aspirated and unaspirated tokens at three points in time (10, 30, and 50 ms after the release). Figure 3 shows these spectra averaged over all tokens of all speakers. Although the formant peaks in these grand average spectra are somewhat flattened because of between-speaker variability in absolute formant frequencies, the general pattern is fairly representative of individual speakers' utterances. Three aspects deserve attention. First, the upward shift in $F_2$ during aspiration is evident, except in the first time frame for alveolar stops, where the spectrum reflects the [s]-like frication noise that is part of the release burst (cf. Figure 1). The $F_2$ peak is rather broad for [p$^h$a], which was also true for most individual speakers' spectra. On its lower skirt, a raised and attenuated $F_1$ (see Kallail & Emanuel, 1984a, 1984b) may have contributed to this prominence. On the upper skirt, additional subglottal resonances may have occurred (Fant, Ishizaka, Lindqvist, & Sundberg, 1972), although we did not observe any distinct peaks in individual spectra that could be identified with such resonances.

Second, it is obvious that the spectrum during aspiration has a different tilt from that during voicing. Acoustic theory predicts a -12 dB/octave spectral slope when the source is voiced, and a -6 dB/octave slope when the source is noise from the glottis (Fant, 1960; Hillman, Oesterle, & Feth, 1983). Although the spectra in Figure 3 are plotted on a linear frequency scale and include high-frequency pre-emphasis of approximately 6 dB/octave above 1 kHz, it is clear that they roughly conform to the predictions. If a correction for pre-emphasis were applied, all spectra would have a downward tilt, the voiced spectra more so than the aspirated ones, as predicted. Labial and alveolar tokens do not seem to differ in spectral tilt.

Third, the relative amplitude of aspiration should be noted. It is especially difficult to locate information in the literature on this parameter, which is often a source of frustration in synthesizing aspirated stops. As can be seen, the levels of voiced and aspirated spectra converge between 3.5 and 4 kHz but diverge increasingly at lower frequencies. The differences observed are somewhat larger than predicted on the basis of a 6 dB/octave slope difference; in fact, they are more in accord with a linear 6 dB/kHz slope difference (cf. Hillman et al., 1983): On the average, the levels of voiced and aspirated $F_3$ peaks differed by 11 dB, and those of $F_2$ peaks by 18 dB, with very similar differences for labials and alveolars. Level differences were even larger in the $F_1$ region, due to the reduction of $F_1$ during aspiration. There was enormous individual variability, however, in the absolute magnitude of these differences: $F_3$ level differences ranged from 4 to 17 dB across speakers, and $F_2$ level differences from 7 to 27 dB, probably reflecting individual differences in source spectra.

## Summary and Conclusions

We have shown that aspirated labial and alveolar stop consonants preceding [*a*] have $F_2$ transitions that start at significantly higher frequencies than those of unaspirated cognates. The difference gets smaller over time and disappears with voice onset, which suggests that it is due to upward shifts in vocal tract resonances caused by the open (and gradually closing) glottis during aspiration. These data replicate and extend earlier observations by others, and they provide a valuable guideline for improved speech synthesis. Fant (1973, p. 131) recommended long ago that a "minor correction for the effect of glottal opening on the F-pattern" be added in synthesis, and noted that "an open glottis increases $F_2$ and $F_3$ by about 50-100 Hz." Our data suggest that, in the context of [*a*], the effect is about twice as large but restricted to $F_2$. It is astonishing that this difference has gone relatively unnoticed for so long, and that it has been completely ignored in the long series of studies employing synthetic stop-consonant-vowel (mostly [-*a*] or [-æ]) syllables and VOT continua over the last 20 years.

For reasons that are not well understood, the raising of $F_2$ during aspiration seems to be absent for high front vowels such as [i] (Gay, 1978; Kallail & Emanuel, 1984a, 1984b). It might be predicted, then, that the perceptual category boundaries on [pi]-[ti] and [p$^h$i]-[t$^h$i] continua should be similar. Unfortunately, this interesting prediction is not testable because $F_2$ transitions do not reliably differentiate labial and alveolar stops in [i] context (see, e.g., Kewley-Port, 1982). Another prediction more amenable to test is that, unless there is differential coarticulation (Fant, 1973), the $F_2$ transitions of whispered [p*a*] and [t*a*] (i.e., intended /b*a*/ and /d*a*/) should not differ from those of [p$^h$*a*] and [t$^h$*a*], and the category boundary on a noise-excited synthetic labial-alveolar continuum should likewise be similar to that on a [p$^h$*a*]-[t$^h$*a*] continuum.

The difference in $F_2$ onset frequencies between aspirated and unaspirated stops preceding [*a*] provides a sufficient explanation of the reliable perceptual shift in the labial-alveolar category boundary on a formant transition continuum as a function of VOT. The magnitude of perceptual boundary shifts reported in the literature (expressed in terms of $F_2$ onset frequency, about 100 Hz on the average) matches the magnitude of the average acoustic difference in $F_2$ transitions. If aspiration is introduced in a synthetic syllable without changing the $F_2$ transition, as has been the custom, listeners *expect* the transition to be higher and therefore perceive the stimulus as relatively more labial. The effect of glottal opening on vocal tract resonances thus seems to be represented in listeners' tacit knowledge of phonetic regularities. Even though the boundary shift is essentially an artifact of primitive synthesis methods, it serves to remind us of the rich store of phonetic knowledge that listeners refer to in speech classification. Identification of speech depends as much on what listeners know about the sounds and gestures of their language as on what is in the acoustic signal (cf. Repp, 1987).

## References

Alfonso, P., & Daniloff, R. (1980). Allophonic backward masking of stop consonants. *Phonetica*, *37*, 355-376.

Catford, J. C. (1977). Fundamental problems in phonetics. Bloomington, IN: Indiana University Press.

Danaher, E. M., & Pickett, J. M. (1975). Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss. *Journal of Speech and Hearing Research*, *18*, 261-271.

Darwin, C. J., & Seton, J. (1983). Perceptual cues to the onset of voiced excitation in aspirated initial stops. *Journal of the Acoustical Society of America*, *74*, 1126-1135.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, *27*, 769-773.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fant, G. (1973). Stops in CV-syllables. In *Speech sounds and features* (pp. 110-142). Cambridge, MA: MIT Press.

Fant, G., Ishizaka, K., Lindqvist, J., & Sundberg, J. (1972). Subglottal formants. *Speech Transmission Laboratory Quarterly Progress and Status Report*, *1*, 1-12. Stockholm: Speech Transmission Laboratory, Royal Institute of Technology.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, *63*, 223-230.

Grunke, M. E., & Pisoni, D. B. (1982). Some experiments on perceptual learning of mirror-image acoustic patterns. *Perception & Psychophysics*, *31*, 210-218.

Haggard, M. P. (1970). The use of voicing information. *Speech Synthesis and Perception*, *2*, 1-14. England: University of Cambridge.

Hannley, M., & Dorman, M. F. (1983). Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss. *Journal of the Acoustical Society of America*, *74*, 40-51.

Hillman, R. E., Oesterle, E., & Feth, L. L. (1983). Characteristics of the glottal turbulent noise source. *Journal of the Acoustical Society of America*, *74*, 691-694.

Hirose, H. (1977). Laryngeal adjustments in consonant production. *Phonetica*, *34*, 289-294.

Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates, and fricatives. *Journal of Phonetics*, *2*, 161-180.

Kallail, K. J., & Emanuel, F. W. (1984a). Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects. *Journal of Speech and Hearing Research*, *27*, 245-251.

Kallail, K. J., & Emanuel, F. W. (1984b). An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects. *Journal of Phonetics*, *12*, 175-186.

Kewley-Port, D. (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *Journal of the Acoustical Society of America, 72*, 379-389.

Lehiste, I. (1964). *Acoustical characteristics of selected English consonants*. Publication 34. Bloomington, IN: Indiana University Research Center in Anthropology, Folklore, and Linguistics.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, I., & Peterson, G. E. (1961). Transitions, glides, and diphthongs. *Journal of the Acoustical Society of America, 33*, 268-277.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech, 1*, 153-167.

Lisker, L., & Abramson, A. (1967). Some effects of context on voice onset time in English stops. *Language and Speech, 10*, 1-28.

Massaro, D. W., & Oden, G. C. (1980) Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America, 67*, 996-1013.

Miller, J. L. (1977). Nonindependence of feature processing in initial consonants. *Journal of Speech and Hearing Research, 20*, 519-528.

Nusbaum, H. C., Schwab, E. C., & Sawusch, J. R. (1983). The role of 'chirp' identification in duplex perception. *Perception & Psychophysics, 33*, 323-332.

Oden, G. C., & Massaro, D. W. (1978) Integration of featural information in speech perception. *Psychological Review, 85*, 172-191.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America, 75*, 224-230.

Ohde, R. N., & Stevens, K. N. (1983). Effect of burst amplitude on the perception of stop consonant place of articulation. *Journal of the Acoustical Society of America, 74*, 706-714.

Öhman, S. E. G. (1965). On the coordination of articulatory and phonatory activity in the production of Swedish tonal accents. *Speech Transmission Laboratory Quarterly Progress and Status Report, 2*, 14-19. Stockholm: Royal Institute of Technology.

Perl, N. (1975). Masking versus hemispheric sharing in the processing of split-formant stimuli. *Speech Perception 2*, No. 4, 31-37. Belfast: Department of Psychology, Queen's University.

Perl, N., & Haggard, M. (1974). Masking versus hemispheric sharing of processing for speech sounds. *Speech Perception 2*, No. 3, 47-54. Belfast: Department of Psychology, Queen's University.

Rand, T. C. (1974). Dichotic release from masking for speech. *Journal of the Acoustical Society of America, 55*, 678-680.

Remez, R. E. (1987). Neural models of speech perception: A case history. In S. Harnad (Ed.), *Categorical perception*. New York: Cambridge University Press.

Repp, B. H. (1978). Interdependence of voicing and place decisions for stop consonants in initial position. *Haskins Laboratories Status Report on Speech Research, SR-53* (vol. 2), 117-150.

Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech, 27*, 173-189.

Repp, B. H. (1984). Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. *Language and Speech, 27*, 245-254.

Repp, B. H. (1987). The role of psychophysics in understanding speech perception. In M. E. H. Schouten (Ed.), *The Psychophysics of Speech Perception*. The Hague: Martinus Nijhoff.

Sawusch, J. R., & Pisoni, D. B. (1974). On the identification of place and voicing features in synthetic stop consonants. *Journal of Phonetics, 2*, 181-194.

Schwab, E. C. (1981). *Auditory and phonetic processing for tone analogs of speech*. Unpublished Ph.D. dissertation, State University of New York at Buffalo.

Smith, P. T. (1973). Feature-testing models and their application to perception and memory for speech. *Quarterly Journal of Experimental Psychology*, *25*, 511-534.

Turek, S. V., Dorman, M. F., Franks, J. R., & Summerfield, Q. (1980). Identification of synthetic /bdg/ by hearing-impaired listeners under monotic and dichotic formant presentation. *Journal of the Acoustical Society of America*, *67*, 1031-1040.

Zue, V. W. (1976). *Acoustic characteristics of stop consonants: A controlled study*. Technical Report No. 523. Lexington, MA: Lincoln Laboratory.

# SENSITIVITY TO INFLECTIONAL MORPHOLOGY IN AGRAMMATISM: INVESTIGATION OF A HIGHLY-INFLECTED LANGUAGE*

**K. Lukatela,† S. Crain,† and D. Shankweiler†**

**Abstract.** *We present the results of a study with six Serbo-Croatian speaking agrammatic patients on a test of inflectional morphology in which subjects judged whether spoken sentences were grammatical or ungrammatical. Sensitivity to two kinds of syntactic features was investigated in these aphasic patients: 1) subcategorization rules for transitive verbs (which must be followed by a noun in the accusative case; intransitive verbs can be followed by nouns in other noun cases); 2) sensitivity to the inflectional morphology marking noun case. The test items consisted of three-word sentences (noun-verb-noun) in which verb transitivity and appropriateness of the case inflection of the following noun were manipulated. Results of the grammaticality judgment task show that both syntactic properties are preserved in these patients.*

## INTRODUCTION

Recent research on Broca-type aphasia has suggested that syntactic deficits in speech production have parallels in speech comprehension. It has been argued that Broca patients with agrammatic output not only tend to omit many grammatical words and grammatical morphemes in their productions, but also fail to process these words properly in comprehension, although special tests were required to bring these problems to light. An important claim in this regard was made by Bradley, Garrett, and Zurif (1980), who offered a unified account of Broca-type aphasia encompassing both production and comprehension, based on results obtained using lexical decision and picture verification tasks.

---

* *Brain and Language*, in press.

† Also University of Connecticut

*Acknowledgment*. We thank Nina Mihajlović of the Clinic for Neurophysiology and Pathology in Belgrade, and Gordana Ocić of the Neurological Clinic of the Medical Faculty, University of Belgrade, who permitted us to test their patients, and provided valuable clinical information. Thanks are also due to Leonard Katz for help with analysis of the data. Portions of this research were supported by a Program Project Grant to Haskins Laboratories from the National Institute of Child Health and Human Development (HD-01994).

However, using a different experimental task, other researchers have found retained capacities of agrammatic aphasics to apprehend syntactic structures and to process the same closed-class items that are so often absent in their speech. Retained ability of English-speaking agrammatics to detect a variety of syntactic anomalies was uncovered by Linebarger, Schwartz, and Saffran (1983) using a grammaticality judgment task. They found that so-called agrammatics perform at much better than chance level in judging the acceptability of many syntactic structures, including ones that hinge on the availability of closed-class items (e.g., auxiliaries). Such evidence is clearly incompatible with any hypothesis that tries to explain agrammatism as loss of tacit knowledge necessary to compute syntactic structure.

Subsequent work by Crain, Shankweiler, and Tuller (1984) supported and extended the finding of preserved receptive processes in the context of severely limited production. Their agrammatic subjects showed retained ability to detect anomalies involving prepositions, determiners, particles, and auxiliary verbs—closed-class items that are often missing in the productions of Broca-type aphasics. Moreover, the agrammatic subjects in this study were pressed to make judgments of grammaticality "on-line," a maneuver that forestalls the possibility that they might be adopting procedures for judging grammaticality that do not appropriately reflect their normal syntactic parsing routines.

The present study pursues the issue of receptive capabilities in agrammatism in patients who speak a language quite unlike English. If it is correct to characterize agrammatism in linguistic terms, then losses in language function that follow lesions in specific language zones will occur across all languages, making agrammatism a universal phenomenon. Still, the particular effects of lesions may vary with structural differences among languages, because languages sometimes employ different means to achieve the same grammatical ends. Thus the same neurological deficit could produce different patterns of symptoms in speakers of different languages. Naturally, the variation in expression of aphasia caused by cross-language differences cannot be without limit if grammatical devices are expressions of a Universal Grammar and subject to its constraints (Chomsky, 1981).

These considerations underscore the importance of cross-language studies of aphasia in evaluating theoretical hypotheses about the source of agrammatism. Among the criteria of theoretical adequacy is the requirement that we should be able to predict and account for the manifestations of agrammatism in different languages.

A recent account of agrammatism proposed by Grodzinsky (1984) gives due weight to such cross-linguistic considerations, and, indeed, makes detailed predictions about the manifestations of agrammatism in several languages. On his account, different languages will have associated with them different patterns of impairment, with the patterns reflecting a common principle: misselection of closed-class words (i.e., the class that includes articles, auxiliary verbs, particles, and prepositions) within the same syntactic category. Other explicit predictions are made, including the prediction (i) that closed-class items will not be missing entirely in all languages, and (ii) distinctions between closed class items belonging to different syntactic categories should be preserved despite the loss of sensitivity to distinctions within a category.

As to the first point, Grodzinsky's theory contrasts free-standing grammatical morphemes, which are often missing entirely in the productions of English-speaking agrammatics, with bound morphemes (grammatical affixes). According to the theory, inflectional affixes will be neglected by agrammatics only when they are unessential to the "well-formedness" of the lexical item—if, in other words, the lexical item without the affix maintains its status as a word. The second prediction of the theory, that between-class sensitivity is preserved in agrammatism, follows from the proposal that what is lost in agrammatism is the lexical content normally present at the terminal nodes of closed-class categories. Information about "part-of-speech" is available, but the particular words are not.

The present study is designed to investigate Grodzinsky's hypotheses, taking advantage of a cross-language difference in use of closed-class morphology. Languages that have few word order constraints are usually also highly inflected; they make heavy use of bound morphemes. On the other hand, fixed word-order languages commonly use word order to mark the same grammatical phenomena that are handled by inflectional morphology in nonconfigurational languages.

Pursuing this distinction, we note that in English the order of constituents is a fundamental device for indicating both semantic and syntactic relationships. German and Serbo-Croatian, in contrast to English, are relatively free-word-order languages. In Serbo-Croatian, morphological inflection is used to express grammatical relations that are expressed by word order in English. Unlike English, where case is conveyed either by word order (or by a free standing preposition or pronoun), Serbo-Croatian marks case relations by noun inflections, and imposes comparatively few restrictions on word order. In order to construct a grammatically correct structure, words have to match in gender, number, person, and noun case. This is accomplished by adding an appropriate suffix, an inflectional morpheme, to the word stem. The fact that the morphology of closed-class items plays such an important role in Serbo-Croatian makes it an ideal language to contrast with English, in testing detailed theoretical claims like Grodzinsky's.

Previous research has shown that both German-speaking and Serbo-Croatian-speaking agrammatics show some degree of sensitivity to case inflection even when the test sentence departs from standard word order (Friederici, 1982; Heeschen, 1980; Smith & Bates, 1985; Smith & Mimica, 1984). Heeschen found that German Broca's aphasics were in error 18% of the time in matching semantically reversible sentences to pictures when standard word order was presented, and in error 27% of the time when standard word order was violated. In an object-manipulation study with Serbo-Croat aphasics, Smith and Mimica showed that agrammatics are differentially sensitive to three types of cues: closed-class morphology, semantic constraints, and word order. Agrammatics were impaired relative to normals when forced to rely on case inflection cues alone. However, it was found that sentence understanding in agrammatic users of Serbo-Croatian was facilitated by a convergence of cues that, in combination, often led to successful processing of sentences. The available data on agrammatism in different languages neither confirm nor disconfirm Grodzinsky's hypothesis that within-class sensitivity to bound morphemes should be impaired in agrammatics who speak a language with relatively free word order. Some impairment is evident, but in the use of convergent cues to assign noun case, there is also evidence of some sparing of function that does not accord well with Grodzinsky's account.

Problems associated with the choice of task to assess grammatical competence merit comment. The findings we have just discussed indicate that aphasic subjects perform better on some tasks than others. Tasks that minimize extraneous demands, e.g., the grammaticality judgment task, have proven more successful in uncovering retained syntactic ability than tasks like picture verification and object manipulation. The latter have been found to underestimate the extent of agrammatics' competence. Consequently, in much previous research, failures of agrammatics to use closed-class morphological items in analysis of sentences may have reflected a processing limitation, and not a structural deficit per se. For these reasons it seems to us that past research does not provide the data needed for a definitive test of Grodzinsky's specific claims about the linguistic source of agrammatic comprehension errors.

The present study focuses specifically on the processing of bound morphemes marking noun case by Yugoslavian agrammatics who were native speakers of Serbo-Croatian. We chose to use elicited grammaticality judgments as the task in order to avoid introducing extraneous processing factors that would otherwise be confounded with syntactic parsing in object-manipulation and picture-verification tasks. The Serbo-Croatian-speaking agrammatic aphasics were tested for retained sensitivity to noun inflections in the context of the contrast between transitive and intransitive verbs. This maneuver allowed us to test Grodzinsky's hypothesis that distinctions within the same closed-class category should be lost in agrammatism.

In the Serbo-Croatian language, subcategorization is related not only to the meaning but also to the syntactic structure of a noun. Both transitive and intransitive verbs can be directly followed by a noun phrase. If the verb is transitive, however, it must be followed by a noun in the *accusative case*. This feature of Serbo-Croatian offers the opportunity to create transitive vs. intransitive sentences that are minimal pairs. Sentences of both types can be constructed so as to be identical except for the terminal noun suffix. This suffix alone may differentiate a transitive from an intransitive sentence. In English it is impossible to create such minimal pairs because an intransitive verb in English *cannot* be directly followed by a noun phrase, whereas a transitive verb *must* be. (These differences between the two languages are diagrammed schematically in Figure 1.) But of course these differences in subcategorization in English, but not in Serbo-Croatian, necessitate differences in prosody and length.

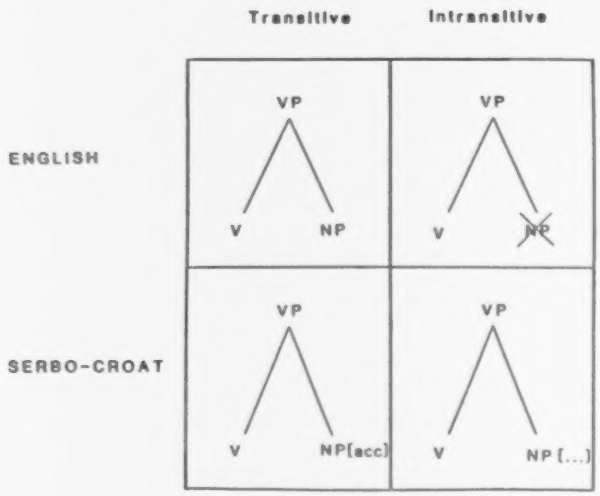

**Figure 1.** The diagram compares the form of the verb phrase for transitive, and intransitive verbs in English, and Serbo-Croatian. Note that in Serbo-Croatian, unlike in English, a noun may follow the verb directly for either a transitive or an intransitive verb. Intransitivity is marked by some other case than the nominative or accusative.

Some evidence has already been obtained, using the grammaticality judgment task, that English-speaking agrammatics are sensitive to the kind of strict subcategorization information that is conveyed by transitive vs. intransitive verbs. However, if Grodzinsky's hypothesis is correct, then Serbo-Croat agrammatics, unlike English-speaking agrammatics, should *not* be sensitive to this subcategorization property of verbs. This is expected because, in Serbo-Croatian, transitivity is captured by affixation and not by word order. Accordingly, Serbo-Croat aphasics should be unable to tell whether there is agreement between a specific verb and the case inflection of a following noun. This is just the kind of cross-language difference that is expected, on Grodzinsky's account, if agrammatism has a linguistic basis.

The ability of Serbo-Croat agrammatics to use subcategorization information was tested by manipulating the case endings of nouns that follow either transitive or intransitive verbs. We wanted to discover whether the subcategorization facts associated with transitive verbs are more accessible to agrammatic aphasics than those associated with intransitive verbs for grammaticality decisions that turn on noun case. Clearly, Grodzinsky's hypothesis would predict that the two classes of verbs should be treated in the same way, so that performance on judgments of grammaticality would be roughly at chance for each. This question was put to an empirical test in our experiment.

To summarize, a much debated issue in neurolinguistics is whether the syntactic deficits of agrammatics in speech production have parallels in speech comprehension. The hypothesis implies that there is some central syntactic processing component that is impaired in agrammatism, and that it is a cause of both comprehension and production difficulties. Our research addresses this issue by focusing on receptive processes in agrammatism from a cross-language point of view. The study had two purposes: first, to identify universal, cross-language characteristics of agrammatism and second, to exploit special characteristics of the Serbo-Croatian language in order to test Grodzinsky's challenging hypothesis that distinctions within the same closed-class category should be lost in agrammatism.

## Subjects

The subjects, who ranged in age from 30-57 years, were six nonfluent aphasics, two females, and four males, all right handed. Their characteristics are summarized in Table 1. In each case, the lesion was confined to the left hemisphere. All had completed at least secondary school. All were outpatients of the Clinic for Neurophysiology and Speech Pathology in Belgrade, Yugoslavia.

Four patients carried the diagnosis of stroke, one was a victim of traumatic insult, and one had a surgically removed tumor. Time since onset of the disorder was at least six months. Three patients (B.S., D.M., and R.N.) were initially mute. CT scans, available for the trauma patient, the tumor patient, and one stroke patient, revealed a lesion predominantly in the inferior posterior region of the left frontal lobe. In addition to the general neurological examination, diagnostic criteria included performance on the Boston Diagnostic Aphasia Examination, which was translated and adapted for Serbo-Croat speakers. Comprehension was relatively good in social contexts, but, as may be seen from the BDAE scores (Table 2), each subject had significant impairment in comprehension.

**Table 1**

**Characteristics of the Aphasic Subjects**

| Subject | Age | Sex | Education | Etiology | Time post-onset |
|---|---|---|---|---|---|
| B. S. | 31 | M | 14 | Trauma | 3 Years |
| D. K. | 36 | M | 14 | CVA | 6 Months |
| D. M. | 33 | F | 16 | Tumor | 4 Years |
| R. N. | 57 | F | 16 | CVA | 3 Years |
| S. P. | 53 | M | 16 | CVA | 5 Years |
| N. M. | 57 | M | 16 | CVA | 1 Year |

In their speech production, all the subjects demonstrated severe-to-moderate agrammatic speech. That is, their speech was effortful, dysprosodic, and telegraphic. Each of the patients made notable production errors on case endings, often using the nominative case in linguistic contexts in which this case was inappropriate. However, none of these errors resulted in nonwords. Representative examples of speech production are given in Table 2.

A group of normal subjects, matched in age and education was also included in the experiment.

## Materials

The experimental materials consisted of 64 grammatical and 64 ungrammatical sentences, each containing three words (noun-verb-noun). Half of the grammatical sentences incorporated a transitive verb followed by the accusative object noun and half incorporated an intransitive verb followed by an adverbial noun, usually in the instrumental case. All the words in the sentence were balanced for length and frequency of occurrence. By varying transitivity, four forms of each sentence were generated as shown in these examples:

1) Seljak obradjuje polje.
(The farmer is cultivating the field.)

2) Seljak trci poljem.
(The farmer is running through the field.)

*3) Seljak obradjuje poljem.
(The farmer is cultivating through the field.)

*4) Seljak trci polje.
(The farmer is running the field.)

## Table 2

### Performance on Selected Portions of the BDAE

| **Subject** | **BDAE Comprehension** | | | | **Speech Production*** |
|---|---|---|---|---|---|
| | **A** | **B** | **C** | **D** | |
| B. S. | 16/20 | 4/15 | 9/12 | 8/10 | Pa.. mama brise tanjir. De...decko..kolaci.. devojcica uzmi uz.. uzima.. Voda curi...<br>Well.. mama is drying the plate. The b..boy ...cookies...the girl take ta...is talking... The water is leaking. |
| D. K. | 20/20 | 5/15 | 6/12 | 6/10 | Voda. Devojcica. Sudove pere. Voda tece. Dete i devojcica se...Ne znam da kazem. Kolace. Devojcica se oklize i pala. Gotovo.<br>The water. The girl. Washing the dishes. Water is leaking. The kid and the girl...I don't know how to say. Cookies. The girl is slipping and fallen. End. |
| D. M. | 18.5/20 | 12/15 | 8/12 | 7/10 | Brat i sestra. Hoce kolace. Mama pere. Ta..tanjir. Voda .. Ne mogu da kazem.<br>Brother and the sister. Want the cookies. Mama is washing. The pla...plate. Water. I can not say. |
| R. N. | 19/20 | 10/15 | 6/12 | 2/10 | Mama i tata...ne, brise sudove. Ne mogu da kazem. Ne mogu da kazem. Vidi kako ovde drzi...Ne mogu da kazem.<br>Mama and daddy...no, drying dishes. I can not say. I can not say. Look how is holding.. I can not say. |
| S. P. | 18.5/20 | 14/15 | 6/12 | 7/10 | Kujna. Mama pere...ovaj tanjir. A ovaj decak i devojcica. To je...Daje sestri kolace. Ova je voda pri..pri..E, voda je pr-li-la. Voda je prelila u sudoperu. Solja.<br>Kitchen. Mama is washing...this plate. boy and girl. This is...Is giving cookies to the sister. This water...is li...li... Water is lea-king. Water is leaking into the sink. The cup. |
| N. M. | 12.5/20 | 8/15 | 6/12 | 6/10 | Ovaj...dete je ustalo da pojede pekmez a ova zena je prosula vodu sto je htela da pere. Pa je sve oprala.<br>This...child got up to eat the jam and this woman has spilled the water cause she wanted to wash. She washed everything. |

A - Body Part Identification.
B - Commands.
C - Complex Ideational Material
D - Reading Sentences and Paragraphs
*Patient's description of "Cookie theft" picture, BDAE.

It will be noted that in each of the above sentences, the correct grammatical form depends on just the last (unstressed) phoneme of the last word in the sentence. It should be noted also that some of the critical nouns preserve their lexical well-formedness even when they appear in unmarked forms (i.e., nominative and accusative).

## Design and Procedure

The sentences were tape recorded and systematically distributed in four groups. Each sentence was read once, with normal speed and intonation. Ungrammatical sentences were read with the intonation appropriate for the corresponding grammatical sentence (i.e., with a correct case inflection). The subjects listened to the sentences over headphones. Their task was to indicate for each sentence whether it was grammatically correct or not. The subjects responded by pressing one of two keys, marked YES and NO. Each subject participated in four individual sessions, one session per week for four consecutive weeks. Each new session started with ten practice sentences to familiarize the subject with the task.

## Results

First, we present an analysis of the error data by subject. Percent of errors for each sentence type is given in Table 3.

Table 3

Percent of Errors for Aphasic Subjects by Sentence Type

| | Grammatical sentences | | Ungrammatical sentences | |
|---|---|---|---|---|
| Subject | Transitive verb | Intransitive verb | Transitive verb | Intransitive verb |
| B.S. | 10.8 | 14.0 | 16.0 | 24.0 |
| D.K. | 4.0 | 6.0 | 4.0 | 10.8 |
| D.M. | 0.0 | 8.0 | 12.0 | 6.4 |
| R.N. | 10.8 | 10.8 | 9.2 | 14.0 |
| S.P. | 6.0 | 9.2 | 14.0 | 16.0 |
| N.M. | 1.6 | 4.0 | 8.0 | 10.8 |
| Mean | 5.5 | 8.7 | 10.5 | 13.7 |

The table shows that the error percentage scores of the individual subjects co-varied with the severity of their aphasia, as measured by neurologists' ratings. It is important to note, however, that all of the subjects were well above chance level in responding correctly to the inflections of the terminal word in the target sentences. The same pattern of errors is apparent for all subjects despite differences in etiology, age, and severity.

Also shown in Table 3 is the analysis of errors by sentence type. Sentences of Type 4 evoked the most errors (i.e., grammatically incorrect sentences with an intransitive verb), and those of Type 1 evoked the fewest errors (i.e., grammatically correct sentences with a transitive verb).

The error data were subjected to analysis of variance by subjects and by items, comparing the factors of group, grammaticality, and transitivity. In both the analyses by subjects and by items there was a significant effect of grammaticality, $F1(1,5) = 16.74, p < 0.01; F2(1,31) = 8.73, p < 0.01$. This means that grammatically correct and grammatically incorrect sentences were not equally difficult for the subjects. It proved to be easier for the subjects to give a correct judgment when the correct inflections were presented.

Analysis of the false alarms indicates that this effect is not due to the tendency for Broca-type aphasics to be "over-accepting." The fact that they correctly rejected ungrammatical sentences 88% of the time is clear evidence of their retained sensitivity to the closed-class morphology, both in accepting grammatically correct sentences and in rejecting ungrammatical sentences.

The effect of transitivity was also significant both by subjects and by items: $F1(1,5) = 10.00, p < 0.025; F2(1,31) = 7.41, p < 0.01$. This indicates that these agrammatic subjects were sensitive to subcategorization requirements that, as we saw, require them to attend to noun inflections. We interpret this result to mean that Broca-type aphasics have preserved information in their lexicons about the complements of verbs, retaining whether or not they obligatorily require a direct object. Presumably, such stored information serves to "prime" the correct noun inflections by generating a syntactic expectancy for a particular case ending.

A comparison of the accuracy of judgments by aphasic patients with those of control subjects demonstrated that although the patient's performance was relatively successful, it was clearly depressed compared to the nearly error-free performance of control subjects. Detection of ungrammatical sentences occurred with an average accuracy of 99.2%, whereas grammatical sentences were correctly identified 98.6% of the time.

An interesting post-hoc observation was made concerning the lexical items that preserve their lexical well-formedness even in the unmarked form. It should be noted that for some nouns the unmarked nominative case is identical to the word stem. For these nouns the other case-inflections are simply appended to the stem (as in Table 4, Column 1). These nouns keep their lexical well-formedness even when the case inflections are neglected. For all other nouns (as in Column 2), the nominative form and other case forms are different from the word stem. For the latter class of nouns, the stem needs a case inflection in order to be a word.

Grodzinsky (1984) has proposed that agrammatics should have difficulty processing inflections of the first class of nouns but not the second class. In the case at hand, this hypothesis would predict that aphasics should make more mistakes when they are processing a sentence in which the critical noun-item belongs to the first class of nouns (nouns in the unmarked nominative case). For example, an aphasic subject should reject a grammatically correct sentence in which the critical noun is inflected with some case other than the nominative or accusative case. This would happen if the subjects were capable of processing the noun only by treating it as if it were in the nominative (unmarked) case. On the other hand, whenever the critical noun is in the nominative case, a subject should have a tendency to accept the sentence even if ungrammatical.

**Table 4**

**Case Inflections for Two Classes of Nouns in Serb-Croatian**

| | Class 1 | Class 2 |
|---|---|---|
| Nominative | sto- (table) | farb-a (color) |
| Genitive | sto-la | farb-e |
| Dative | sto-lu | farb-i |
| Accusative | sto- | farb-u |
| Instrumental | sto-lom | farb-om |

However, it was found that nouns in the marked cases were not significantly more difficult than those in the unmarked case for our subjects. This finding disconfirms Grodzinsky's prediction.

## Discussion

The main result was that agrammatics in this study proved to be capable of using bound closed-class morphemes in sentence processing. Each of the six Serbo-Croat-speaking agrammatic patients showed evidence of retained ability to respond selectively to noun inflections marking noun case and verb transitivity. The finding of retained syntactic competence is consistent with earlier findings of Smith and Mimica (1984) in Serbo-Croatian and of Heeschen (1980) in German.

The findings are also consistent with recent work with English-speaking agrammatics who showed a retained ability to perform judgments of sentence grammaticality (Linebarger et al., 1983). Further, the results are consistent with the indications that agrammatic aphasics are capable of carrying out syntactic analyses on line (Crain et al., 1984; Tyler, 1985). The subjects of the present study, like their English-speaking counterparts, demonstrated retained sensitivity to syntactic category even when the category is marked by affixation and not by word order or by free-standing grammatical morphemes.

As noted in the introduction, this result could not have been presupposed. It is conceivable that agrammatics would be capable of exploiting one indicator of syntactic category, but not another. Given the indications that agrammatics are deficient in use of closed-class vocabulary items, one might be led to suppose that some ability to use word order is retained while ability to use the closed class morphology is lost. The structure of English does not permit us to distinguish between these possibilities, because word order and the introduction of prepositions such as *to* and *by* are the only devices available for marking noun case. But, as we noted, the Serbo-Croatian language, by virtue of its rich inflectional system, enables us to test the effect of relying solely on inflectional morphemes for marking case. Both transitive and intransitive verbs can be directly followed by a noun phrase. This feature of Serbo-Croatian made possible the creation of transitive and intransitive sentences that are minimal pairs, differing solely in one noun suffix.

In summary, our Serbo-Croat speaking agrammatic patients showed retained sensitivity to noun inflections marking the transitive/intransitive verb within the context of the sentence judgment task. The error rate was remarkably low, averaging only 12% for aphasic subjects across conditions. The finding of preserved sensitivity to case in this context clearly fails to support Grodzinsky's (1984) hypothesis that distinctions within the same closed-class category should be lost in agrammatism.

The questions we raised about the ability of agrammatics to use closed class morphology in sentence processing were also addressed in the recent research of Smith and Mimica (1984), to which we have referred. In that study, also, Serbo-Croatian-speaking agrammatics showed better than chance ability to use case inflection in the assignment of agent-object relations, but the error rate was much higher than in the present study. The large differences in rate of correct responses may be attributable to the task. Smith and Mimica used an object manipulation task, which is known to impose a considerable burden on short-term memory processes (Hamburger & Crain, 1984).

An explanation of agrammatics' performance failures in terms of processing limitations rather than loss of syntactic competence is fully in line with the other findings of the Smith and Mimica study. These investigators explored the effects on the acting out task of association or dissociation of three variables: word order, animacy, and case inflection. When all three factors were concordant, Broca's aphasics performed with only 10% of error, whereas when two factors were in competition, performance fell to near chance level (42% errors). In their terms, situations of conflict, such as that created by use of a nonpreferred word order, create "cognitive overload."

Taken together, the findings of the present study are consistent with other research both on richly inflected languages and on fixed word order languages like English. The weight of the evidence supports the view that comprehension deficits in agrammatism do not reflect loss of either the knowledge or ability to access members of the closed-class lexicon in extracting the syntactic structure of a sentence. Access to grammatical knowledge is impaired, to be sure, but access can often be attained successfully in circumstances that minimize processing load.

A comparison of agrammatics' performance across tasks shows that subjects who standardly fail in an object manipulation task may succeed in a grammaticality judgment task that tests comprehension of the same linguistic structures. This implies that all necessary syntactic structures may be preserved in the so-called agrammatism of Broca's aphasia and that problems in some other part of the language apparatus are responsible for failures of comprehension. There is increasing support for the view that complex behaviors are products of an interaction between many different and independent subsystems, each performing a unique and special role. In agrammatism, a likely source of comprehension problems is a verbal working memory limitation. There is evidence that the phonological processing system on which the verbal working memory depends is often damaged in the nonfluent aphasias (Caramazza, Berndt, & Basili, 1983; Martin, 1985).

In sum, the findings of the present study are consistent with the main body of research on sentence processing in Broca's aphasia in suggesting that the link between linguistic competence and linguistic performance is not fully preserved. Tacit knowledge of syntax is seen to be intact under circumstances that tax working memory as little as possible. However, linguistic knowledge is less accessible in contexts, including many everyday contexts, that place heavy demands on working memory. Thus, the data we have reviewed implicate disturbances of language subsystems other than the syntactic component and suggest that studies investigating the role of such processing components as working memory will be important in the future.

## References

Bradley D., Garrett M., & Zurif E. (1980). Syntactic deficits in Broca's aphasia. In D. Caplan (Ed.), *Biological studies of mental capacities*. Cambridge, MA: MIT Press.

Caramazza, A., & Zurif, E. (1976). Dissociation of algorithmic and heuristic processes in language comprehension. *Brain and Language, 3*, 572-582.

Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures*. Dordrecht, Holland: Foris Publications.

Crain S., Shankweiler D., & Tuller B. (1984). *Preservation of sensitivity to closed-class items in agrammatism*. Talk presented at the Annual Meeting of the Academy of Aphasia, Los Angeles, October, 1984.

Grodzinsky, Y. (1984). The syntactic characterization of agrammatism. *Cognition, 16*, 99-120.

Hamburger, H., & Crain, S. (1984). Acquisition of cognitive compiling. *Cognition, 17*, 85-136.

Heeschen C. (1980). Strategies of decoding actor-object relations by aphasic patients. *Cortex, 16*, 5-19.

Kean. M. L. (1979). Agrammatism: A phonological deficit? *Cognition, 7*, 69-84.

Kolk, H. J., & Friederici, A. D. (1985). Strategy and impairment in sentence understanding by Broca's and Wernicke's aphasics. *Cortex, 21*, 47-68.

Kolk, H. J., & VanGrusven, M. J. (1984). Metalinguistic judgments on sentence structure in agrammatism: A matter of task misinterpretation. *Neuropsychologia, 22*, 31-39.

Linebarger, M., Schwartz M., & Saffran E. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition, 13*, 361-392.

Martin, R. C. (1985). *The relationship between short-term memory and sentence comprehension deficits in agrammatic and conduction aphasics*. Paper presented at the Annual Meeting of the Academy of Aphasia, Pittsburgh, Pa.

Smith S., & Bates, E. (in press). Accessibility of case and gender contrasts for agent-object assignment in Broca's aphasia and fluent anomics. *Brain and Language*.

Smith S., & Mimica I. (1984). Agrammatism in a case-inflected language: Comprehension of agent-object relations. *Brain and Language, 21*, 274-290.

Tyler, L. (1985). Real-time comprehension processes in agrammatism: A case study. *Brain and Language, 26*, 405-417.

Zurif, E. B., & Grodzinsky, Y. (1983). Sensitivity to grammatical structure in agrammatic aphasics: A reply to Linebarger, Schwartz, and Saffran. *Cognition, 15*, 202-213.

# INTENTIONALITY: A PROBLEM OF MULTIPLE REFERENCE FRAMES, SPECIFICATIONAL INFORMATION, AND EXTRAORDINARY BOUNDARY CONDITIONS ON NATURAL LAW*

**M. T. Turvey†**

It is refreshing to see a scholar who is largely sympathetic to the so-called information processing or representational/computational approach to cognitive systems recognizing its fundamental inadequacies. To be blunt, that approach fails to come to terms with either information or intentionality. Sayre's response to these inadequacies, however, keeps close to the received view. He assumes that a biologically and psychologically relevant sense of information can be provided by the mathematical theory of communication; he assumes that intentionality amounts to representation. These assumptions are bolstered by the closely cognate beliefs that intentionality is to be ascribed to some roughly midway-state in the classical afferent-efferent link and that there is a metamorphosis from meaningless states to meaningful states. To his credit, Sayre aspires to make the representations genuine. He wants them to stand for real things. He wants the transition from meaningless sensory states to meaningful perceptual states to be (mathematically) principled.

From my perspective as a proponent of the ecological approach to perceiving-acting (see Gibson, 1979; Turvey, Shaw, Reed, & Mace, 1981), Sayre's sentiments are right but his premises are wrong. Nor surprisingly, I find his treatment of intentionality disappointing. I concur with Sayre's implicit wish for a concerted effort to *naturalize* (my word) intentionality, but my preference is to keep the deliberations very close to natural science and the search for lawful regularities. Sayre is quite right in his assessment that an attempt to devise an explanation of intentionality in the Turing reductionism/token physicalism perspective of cognitive science (which denigrates intentionality to the states of a computational device) does not have a "ghost of a chance" (Carello, Turvey, Kugler, & Shaw, 1984; Turvey et al., 1981). But he is quite wrong, I believe, in suggesting that pursuing the purer equation of intentionality with representation (relieved of computational procedures) can fare any better.

Intentionality is directedness toward objects. Locomoting terrestrial animals, including humans, direct themselves through openings and around barriers. They direct their limbs in certain ways with respect to a brink in a surface—directing them one way if the brink is where they can step down and another way if it must be negotiated by jumping. Gibson (1966, 1979; Reed &

---

* *The Behavioral and Brain Sciences, 9*, 153-155, 1986. Commentary on Sayre, K. M. (1986). Intentionality and information processing: An alternative model for cognitive science. *The Behavioral and Brain Sciences, 9*, 121-138.

† Also University of Connecticut.

*Acknowledgment.* This work was supported in part by Office of Naval Research Contract N00014-83-K-0083 to Haskins Laboratories.

Jones, 1982) advocated mutually constraining theories of animals and environments (see Alley, in press; Mace, 1977; Michaels & Carello, 1981) as the basis for an understanding of perceiving-acting that addressed such mundane intentional behavior. (This central thesis of the ecological approach, the duality of animal and environment [Shaw & Turvey, 1982], implies that efforts to ground intentionality only in "environmental constraints" will miss the mark. Duality, by the way, is not dualism.) Gibson pursued a perceptual theory that was fundamentally intentional rather than one that is made intentional as an afterthought. With considerable care he identified how an understanding of intentionality of perceiving poses challenges for science on several fronts, and how these challenges might be met. I will describe two of them.

The first challenge is to describe the layout of surfaces with reference to the animal. This move is continuous with the larger lesson of relativity theory: All state descriptions are frame dependent. Reference frames are substantial and are not to be confused with the coordinate systems that abstractly represent them. The properties of an animal to which surface layout must be referred are basically the animal's magnitudes, its morphology, its metabolism. With regard to a brink, the separation of surfaces is in reference to limb magnitudes. Obviously a given brink can be referred to multiple, equally real frames. One frame is the terrestrial frame with distances and durations measured in arbitrary units. This frame is useful to the physicist but it is, by definition, animal-neutral. (In the received view it is mistakenly adopted as the sole objective frame.) Other frames are individual animals. Consequently, the same brink in the terrestrial frame is a place negotiable by leg extension in the frame provided by one (larger) animal not negotiable in this fashion in the frame provided by another (smaller) animal.

A second challenge is to describe how animals can be informed about these frame-dependent environmental properties (affordances) to which their activities are directed. There are two senses in which the term information is used (cf. Turvey & Kugler, 1984). In the indicational/injunctional sense information consists of symbol strings identifying states of affairs ("the situation is so-and-so") or things to be done ("do so-and-so now"). Information in this sense is underconstraining, like a stop sign. The other sense is the specificational sense of Gibson (1979). In the case of vision, information is optical structure *lawfully* generated by facts—properties of surface layout, properties of an animal's movements. This structure does not resemble the facts; rather it is *specific* to them. The ecological argument is that information in the specificational sense meets the above challenge. I will give some examples shortly but I wish to preface them by noting what's at issue in the contrast between the two senses of information.

The indicational/injunctional sense, I believe, fits neatly into a tradition that takes the primary perceptual activity to be discriminating among members of a set and the equilibrium thermodynamics of closed systems as the branch of physics to which discussions of information can be meaningfully referred. In such a system the states are enumerable from the outset. To put it very roughly, the information notion only has to address their individual probabilities, thereby providing a basis for discriminating among them. Living things, however, are open systems. The animal–environment system, in which an animal participates as one of the two mutually tailored components, is open. Significantly, the states of an open system need not be fixed at the outset. Given fluctuations in the microstructure and nonlinearities, a scaling up in one or more variables discontinuously decreases an open system's symmetry. More constraints arise. The system becomes more ordered. New states come into existence. Consequently, the order principle and complexions of Boltzman, and the notion of information that they sustain, are of limited applicability to open physical systems (e.g., Prigogine, 1980), including animal-environment systems.

Open (evolving, developing) systems motivate a different notion of information from closed systems (Kugler, Kelso, & Turvey, 1982; Kugler & Turvey, in press). Sayre makes an offhand remark about the information in the genes and in the phenotype. Efforts to apply classical information theoretic notions to the genotype-phenotype link, conceived as a communication channel, have largely been dismissed. In intuitive terms, the dismissal is based upon a feeling that an information metric should recognize the greater complexity of the full-fledged animal (Waddington, 1968). Even where the open-closed distinction is sidestepped, as in Pattee's (1973, 1977) thoroughgoing and celebrated efforts to detail the problem of a physical interpretation of "genetic information," the conceptions of the mathematical theory of communication have proven to be of little value.

The specificational sense of information is consistent with the perspective that takes perceiving the persisting and changing properties of a thing as primary. For Gibson (1966, 1979) the fundamental question is how to characterize the information that supports the perceiving of *P*; the question of how to characterize the information that supports distinquishing *P* from *Q*, *R*, and so on is secondary and derivative. Suppose that *P* is the animal itself. In locomoting, a terrestrial animal generates forces that displace it relative to the surroundings. There are obvious mechanical regularities to be noted. They are ordinarily expressed through Newton's laws. But this situation also exhibits nonmechanical regularities expressed by non-Newtonian laws of wide (though not universal) scope. For instance, all the densely nested optical solid angles, whose bases are the faces and facets of surfaces and whose apex is the point of observation, change concurrently. An optical flow field—crudely, a smooth velocity vector field—is generated. The global form of the flow, or optical morphology, is specific to the configuration of locomotory forces and to the displacements of the animal. Rectilinear forward locomotion, for example, lawfully generates a dilating parabolic flow; a dilating parabolic flow specifies rectilinear forward locomotion.

This simple but significant example of information in the specificational sense permits me to make briefly some important points that can be more carefully developed (e.g., Solomon, Carello, & Turvey, 1984; Turvey & Carello, 1985, 1986; Turvey et al., 1981). First, optical information in the specificational sense is optical structure whose macroscopic, qualitative properties are nomically dependent upon and specific to (under natural boundary conditions) properties of the animal–environment system. Second, optical information in the specificational sense does not reduce to neural signals in the visual system (see below). Thinking about optical information as alternative (macroscopic, qualitative) descriptions of the photon light field, structured by the layout of material surfaces and defined relative to locations and paths in the transparent medium (air for terrestrial animal), is useful. It aids an understanding of optical information independent of vision and of the kinds of ocular systems that evolved. Optical information in the specificational sense is tied to laws at the ecological scale, laws that relate optical properties to kinetic properties (of the animal-environment system). The ecological approach argues that these laws were the basis for the evolution of, and are the basis for the everyday realization of, locomotor activity and its directedness and intentionality.

Let's extend the example a little. Dilation of an optical solid angle relative to a point of observation specifies the approach of a substantial surface. The inverse of the relative rate of dilation, $\tau$, specifies when the collision will occur if the current kinetic conditions persist

(Lee, 1980). And the rate at which $\tau$ changes has a critical point property below which it specifies that the upcoming collision will be hard (Kugler, Turvey, Carello, & Shaw, 1985; Lee, 1980). The foregoing are not so much quantities as they are local flowfield morphologies and their changes. They specify pending states. They make possible the synchronizing of acts with events–the prospective control of basic behavior. They are meaningful in a very pragmatic sense of the word. Speaking in Dennett's (1983) terms, information in the specificational sense has "intentional features." And to echo Gibson's (1966, 1979) longstanding gripe, the "meaningless to meaningful" problem with which Sayre struggles is not a problem. (Coming to terms with the laws at the ecological scale on which the intentionality of perceiving-acting is founded, and figuring out how to formulate and systematize them, now that's a problem!)

Said succinctly, there is a description of optical structure under which its detection guarantees the intentionality of perceiving. There are other descriptions of optical structure under which it must be translated or processed or interpreted or embellished to *make*perceiving intentional. Sayre is playing with one such description. In this respect it is important to note that Gibson (1966, 1979) avidly denied that optical information in the specificational sense was the sort of thing that could be "processed." It is bizarre, therefore, for Sayre to claim that Marr (1982) is on target with his criticism that Gibson underestimated the complexity of visual information processing. There is a clash of metaphors here. Marr and Sayre are operating in the orthodox metaphor of the nervous system as an efficient cause; for example, it produces percepts. Gibson (1966) sees the nervous system as functioning vicariously in perceiving. It is a part (albeit extremely rich) of the supportive basis for the expression of natural cum ecological laws (cf. Ben-Zeev, 1984). An understanding of the nervous system's role in vision in the support metaphor will be radically different from the processing/producing understanding subscribed to by Marr and Sayre (Kugler & Turvey, in press). At all events, in the ecological view, optical descriptions that invoke processing to render intentionless inputs into intentional percepts are of the wrong kind. They beg too many questions and they cast intentionality as a derivative rather than a primary phenomenon.

The last sentences, of course, are just another way of saying that intentionality should not be reduced to representation. As I remarked above, Sayre's goal of disengaging intentionality from computational procedures is admirable; his insistence on the intentional-representational equation is not. That equation, as I have been trying to stress, diverts us from addressing intentionality in a way that reveals its position in the natural order of things. Consider the following: What are customarily referred to as an animal's or person's intentional contents (cf. Dennett, 1969; Searle, 1983) constitute extraordinary boundary conditions on natural law (especially those laws that are particularly pertinent to the ecological scale). A flying animal aiming to collide gently with a surface will synchronize its deceleration with one value of $\tau$; an acceleration to produce a timely, violent collision will be generated with respect to another value of $\tau$ (e.g., Lee & Reddish, 1981; Lee, Young, Reddish, Lough, & Clayton, 1984; Wagner, 1982). In these simple examples the final conditions—the animal's intentional content–specify the initial conditions that a law (relating optical properties to kinetic conditions) must assume. Examples like this abound, and one of them has been investigated quite thoroughly (Kugler & Turvey, in press). They suggest a profound challenge for naturalizing intentionality: understanding the principles by which intentional contents harness natural laws.

## References

Alley, T. R. (in press). Organism—environment mutuality, epistemics, and the concept of an ecological niche. *Synthese*.

Ben-Zeev, A. (1984). The Kantian revolution in perception. *Journal for Theory of Social Behavior, 14*, 69-84.

Carello, C., Turvey, M. T., Kugler, P. N., & Shaw, R. E. (1984). Inadequacies of the computer metaphor. In M. S. Gazzaniga (Ed.), *Handbook of cognitive neuroscience*. New York: Plenum.

Dennett, D. C. (1969). *Content and consciousness*. Boston: Routledge & Kegan Paul.

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *Behavioral and Brain Sciences, 6*, 343-390.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and coordination*. Chichester: Wiley.

Kugler, P. N., & Turvey, M. T. (in press). *Information, natural law and the self-assembly of rhythmic movements: Theoretical and experimental investigations*. Hillsdale, NJ: Erlbaum.

Kugler, P. N., Turvey, M. T., Carello, C., & Shaw, R. E. (1985). The physics of controlled collisions: A reverie about locomotion. In W. H. Warren, Jr. & R. E. Shaw (Eds.), *Persistence and change: Proceedings of the first international conference on event perception*. Hillsdale, NJ: Erlbaum.

Lee, D. N. (1980). Visuo-motor coordination in space-time. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. New York: North-Holland.

Lee, D. N., & Reddish, P. E. (1981). Plummeting gannets: A paradigm of ecological optics. *Nature, 293*, 293-294.

Lee, D. N., Young, D. S., Reddish, P. E., Lough, S., & Clayton, T. M. H. (1984). Visual timing in hitting an accelerating ball. *Quarterly Journal of Experimental Psychology, 35A*, 333-346.

Mace, W. M. (1977). James J. Gibson's strategy for perceiving: Ask not what's inside your head, but what your head's inside of. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology*. Hillsdale, NJ: Erlbaum.

Marr, D. (1982). *Vision*. New York: W. H. Freeman.

Michaels, C. F., & Carello, C. (1981). *Direct perception*. New York: Prentice-Hall.

Pattee, H. H. (1973). Physical problems of the origin of natural controls. In A. Locker (Ed.), *Biogenesis, evolution, homeostasis*. Heidelberg: Springer-Verlag.

Pattee, H. H. (1977). Dynamic and linguistic modes of complex systems. *International Journal of General Systems, 3*, 259-266.

Prigogine, I. (1980). *From being to becoming: Time and complexity in the physical sciences*. San Francisco: W. H. Freeman.

Reed, E., & Jones, R. (1982). *Reasons for realism: Selected essays of James J. Gibson*. Hillsdale, NJ: Erlbaum.

Searle, J. R. (1983). *Intentionality*. Cambridge: Cambridge University Press.

Shaw, R. E., & Turvey, M. T. (1982). Coalitions as models for ecosystems. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum.

Solomon, J., Carello, C., & Turvey, M. T. (1984). Flow fields: The optical support for skilled activities. In W. F. Straub & J. M. Williams (Eds.), *Cognitive sport psychology.* Lansing, NY: Sport Science Associates.

Turvey, M. T., & Carello, C. (1985). The equation of information and meaning from the perspectives of situation semantics and Gibson's ecological realism. *Linguistics and Philosophy, 8,* 81-90.

Turvey, M. T., & Carello, C. (1986). The ecological view of perceiving-acting: A pictorial essay. *Acta Psychologica, 63,* 133-155.

Turvey, M. T., & Kugler, P. N. (1985). A comment on equation information with symbol strings. *American Journal of Physiology, 246,* R895-896.

Turvey, M. T., Shaw, R. E, Reed, E. S., & Mace, W. M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). *Cognition, 9,* 237-304.

Waddington, C. H. (1968). The basic ideas of biology. In C. H. Waddington (Ed.), *Towards a theoretical biology.* Chicago: Aldine.

Wagner, H. (1982). Flow-field variables trigger landing in flies. *Nature, 297,* 147-148.

# *APPENDIX*

| **Status Report** | | **DTIC** | **ERIC** |
|---|---|---|---|
| SR-21/22 | January - June 1970 | AD 719382 | ED 044-679 |
| SR-23 | July - September 1970 | AD 723586 | ED 052-654 |
| SR-2 | October - December 1970 | AD 727616 | ED 052-653 |
| SR-25/26 | January - June 1971 | AD 730013 | ED 056-560 |
| SR-27 | July - September 1971 | AD 749339 | ED 071-533 |
| SR-28 | October - December 1971 | AD 742140 | ED 061-837 |
| SR-29/30 | January - June 1972 | AD 750001 | ED 071-484 |
| SR-31/32 | July - December 1972 | AD 757954 | ED 077-285 |
| SR-33 | January - March 1973 | AD 762373 | ED 081-263 |
| SR-34 | April - June 1973 | AD 766178 | ED 081-295 |
| SR-35/36 | July - December 1973 | AD 774799 | ED 094-444 |
| SR-37/38 | January - June 1974 | AD 783548 | ED 094-445 |
| SR-39/40 | July - December 1974 | AD A007342 | ED 102-633 |
| SR-41 | January - March 1975 | AD A013325 | ED 109-722 |
| SR-42/43 | April - September 1975 | AD A018369 | ED 117-770 |
| SR-44 | October - December 1975 | AD A023059 | ED 119-273 |
| SR-45/46 | January - June 1976 | AD A026196 | ED 123-678 |
| SR-47 | July - September 1976 | AD A031789 | ED 128-870 |
| SR-48 | October - December 1976 | AD A036735 | ED 135-028 |
| SR-49 | January - March 1977 | AD A041460 | ED 141-864 |
| SR-50 | April - June 1977 | AD A044820 | ED 144-138 |
| SR-51/52 | July - December 1977 | AD A049215 | ED 147-892 |
| SR-53 | January - March 1978 | AD A055853 | ED 155-760 |
| SR-54 | April - June 1978 | AD A067070 | ED 161-096 |
| SR-55/56 | July - December 1978 | AD A065575 | ED 166-757 |
| SR-57 | January - March 1979 | AD A083179 | ED 170-823 |
| SR-58 | April - June 1979 | AD A077663 | ED 178-967 |
| SR-59/60 | July - December 1979 | AD A082034 | ED 181-525 |
| SR-61 | January - March 1980 | AD A085320 | ED 185-636 |
| SR-62 | April - June 1980 | AD A095062 | ED 196-099 |
| SR-63/64 | July - December 1980 | AD A095860 | ED 197-416 |
| SR-65 | January - March 1981 | AD A099958 | ED 201-022 |
| SR-66 | April - June 1981 | AD A105090 | ED 206-038 |
| SR-67/68 | July - December 1981 | AD A111385 | ED 212-010 |
| SR-69 | January - March 1982 | AD A120819 | ED 214-226 |
| SR-70 | April - June 1982 | AD A119426 | ED 219-834 |
| SR-71/72 | July - December 1982 | AD A124596 | ED 225-212 |
| SR-73 | January - March 1983 | AD A129713 | ED 229-816 |
| SR-74/75 | April - September 1983 | AD A136416 | ED 236-753 |
| SR-76 | October - December 1983 | AD A140176 | ED 241-973 |
| SR-77/78 | January - June 1984 | AD A145585 | ED 247-626 |
| SR-79/80 | July - December 1984 | AD A151035 | ED 252-907 |
| SR-81 | January - March 1985 | AD A156294 | ED 257-159 |
| SR-82/83 | April - September 1985 | AD A165084 | ED 266-508 |
| SR-84 | October-December 1985 | AD A168819 | ED 270-831 |
| SR-85 | January-March 1986 | AD A173677 | ED 274-022 |
| SR-86/87 | April-September 1986 | AD A176816 | ED 278-066 |

Information on ordering any of these issues may be found on the following page.

**DTIC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm Corp.
(CMC)
3900 Wheeler Avenue
Alexandria, Virginia 22304-5110

Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Language Behavior Abstracts*, P.O. Box 22206, San Diego, California 92122.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>SR-89/90, 1987 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE *(and Subtitle)*<br>Haskins Laboratories Status Report on Speech Research | | 5. TYPE OF REPORT & PERIOD COVERED<br>Continuation Report<br>January-June, 1987 |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR*(s)*<br>Staff of Haskins Laboratories;<br>Michael Studdert-Kennedy, President | | 8. CONTRACT OR GRANT NUMBER*(s)*<br>HD-01994<br>NO1-HD-5-2910<br>RR-05596<br>NS 18010<br>BNS-8520709<br>NS 13870<br>NS 13617<br>NS 24655 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Haskins Laboratories<br>270 Crown Street<br>New Haven, CT 06511-6695 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>National Institutes of Health<br>National Science Foundation | | 12. REPORT DATE<br>July, 1987 |
| | | 13. NUMBER OF PAGES<br>220 |
| 14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* | | 15. SECURITY CLASS. *(of this Report)*<br>Unclassified |
| | | 15a. DECLASSIFICATION DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT *(of this Report)*

UNLIMITED: Contains no information not freely available to the general public.
It is distributed primarily for library use.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

N/A

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Speech Perception:

Integration, segregation, adaptation, short-term, auditory, prevocalic [m]-[n], contrasts, transitions, second formant, aspirated, unaspirated, stop consonants, [a], vowels, talker-independent information, precedence over nonspeech.

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This report (1 January-30 June) is one of a regular series on the status and progress of studies on the nature of speech, instrumention for its investigation, and practical applications. Manuscripts cover the following topics:

- Integration and segregation in speech perception
- Speech perception takes precedence over nonspeech perception
- Evidence of talker-independent information for vowels
- Controlled variables in sentence intonation

DD FORM 1473 1 Jan 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

**19. Key Words (Continued)**

**Speech Articulation:**
Intonation, sentence, controlled variables, articulatory synthesis, hyperbolic differential equation, solution

**Motor Control:**
Intentionality, reference frames, specificational information, boundary conditions, natural law

**Reading:**
Learning-disabled, derivational morphology, spelling, early ability, word-fragments, cues, crossword, inflectional morphology, aggrammatism, constraint-facilitation, lexical decision, single-words contents, grammatial congruency, violations

**20. Abstract (Continued)**

- Articulatory synthesis: Numerical solution of a hyperbolic differential equation
- Type and number of violations and the grammatical congruency effect in lexical decision
- Low constraint facilitation in lexical decision with single word contexts
- The use of morphological knowledge in spelling derived forms by learning-disabled and normal students
- The development of morphological knowledge in relation to early spelling ability
- The crossword pussle paradigm: The effectiveness of different word fragments as cues for the retrieval of words
- On the possible role of auditory short-term adaptation in perception of the prevocalic [m]-[n] contrast
- Difference in second-formant transitions between aspirated and unaspirated stop consonants preceding [a]
- Sensitivity to inflectional morphology in agrammatism: Investigation of a highly-inflected language
- Intentionality: A problem of multiple reference frames, specificational information, and extraordinary boundary conditions on natural law